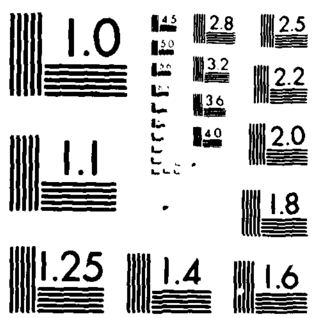
AD-A098 240 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
A GENERALIZED APPROACH TO THE TWO SAMPLE PROBLEM: THE QUANTILE --ETC(U)
APR 81 T J PRIHODA DAAG29-80-C-0070
UNCLASSIFIED TR-B-5 ARO-16992.5-M NL

1 of 2



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

18 ARO 19 16992.5-M

TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843

12

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

LEVEL II

AD A098240

6 A GENERALIZED APPROACH TO THE TWO SAMPLE PROBLEM:

THE QUANTILE APPROACH.

10 Thomas Jeffery Prihoda

Institute of Statistics, Texas A&M University

14 TR-B-5

9 Technical Report, No. B-5

11 Apr[redacted]81

12 138

Texas A & M Research Foundation
Project No. 4226

"Robust Statistical Data Analysis and Modeling"

15 Sponsored by the U.S. Army Research Office
Grant DAAG29-80-C-0070

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

347380

81 4 27 051

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report B-5	2. GOVT ACCESSION NO.: AD-A098240
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A Generalized Approach to the Two Sample Problem: The Quantile Approach	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Thomas Jeffery Prihoda	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0070
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Army Research Office	12. REPORT DATE: April, 1981
	13. NUMBER OF PAGES: 136
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Two Sample Problems, Linear Rank Statistics, Non-parametric Statistics, Location-scale Parameter Estimation, Goodness-of-fit, Quantile Functions, Density Quantile Functions, Data Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See attached sheet

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract

Parzen (1979) suggests a location and scale model for the quantile function (inverse distribution function) of a random variable. We extend this model to the two sample and k-sample problems and some results are given which, when fully implemented, will yield more general solutions in the analysis of variance. Most of the work here concerns the location and scale model suggested by Parzen (1980) for the two sample problem for testing the equality of two distribution functions versus local alternatives.

We implement this model (its tests and estimators) for seven underlying densities. We then provide criteria for choosing or determining whether an underlying density models the differences of the two samples adequately. These criteria allow one to choose the best of several underlying densities for the data. We illustrate these techniques by analyzing data sets from the literature and making comparisons with other authors' techniques. We also show how the Parzen (1980) model is related to many of the techniques developed for studying differences of two samples over the past 50 years. We suggest extensions of Parzen's model. Finally, we give a few simulated examples and suggest what type of simulation study is needed to further define the usefulness of the various models presented in the dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 The Problem

A fundamental problem of statistical theory and application is the two sample problem, i.e., comparing two populations given random samples from each. For example, researchers are often interested in inferring the effect of a treatment on a response variable for some general population. The inference is based on the observed responses from a control group and a treatment group selected from the population being studied. The various methods of dealing with these two data sets have been generalized to k groups and have also been used in developing general statistical theory. The two sample problem has indeed been a cornerstone of statistical science. We begin by giving a series of definitions basic to the approach we shall use. In these definitions we follow Parzen (1961, 1967, 1979, 1980).

1.1.1 Definitions and Notations

We have independent realizations $\{X_1, X_2, \ldots, X_m\}$ and $\{Y_1, Y_2, \ldots, Y_n\}$ of continuous random variables X and Y having continuous increasing distribution functions F(x) and G(x) respectively. The distribution functions F(x) and G(x) often represent those of the control and treatment groups respectively. A popular model for X_i and Y_j, which is assumed in this work, is that both distributions are a location and scale change from a common distribution function, $F_o(x)$, i.e., for $-\infty < x < \infty$,

---

This dissertation will follow the style of the Journal of the American Statistical Association.

$$F(x) = F_o\left(\frac{x-\mu_1}{\sigma_1}\right) \;,\; G(x) = F_o\left(\frac{x-\mu_2}{\sigma_2}\right) \quad .$$

The sample distribution functions are defined by:

$$\tilde{F}(x) = \frac{1}{m} \sum_{i=1}^{m} I(X_i \le x) \;,\; -\infty < x < \infty \;,$$

$$\tilde{G}(x) = \frac{1}{n} \sum_{j=1}^{n} I(Y_j \le x) \;,\; -\infty < x < \infty \;, \tag{1.1}$$

where

$$I(u_i \le u) = 0 \;,\; \text{if } u_i > u \;,$$
$$= 1 \;,\; \text{if } u_i \le u \;.$$

The combined sample distribution function is given by

$$\tilde{H}(x) = \lambda\, \tilde{F}(x) + (1-\lambda)\, \tilde{G}(x) \quad , \tag{1.2}$$

where $\lambda = \frac{m}{N}$ and $N = m+n$. We can regard $\tilde{H}$ as a nonparametric estimator of the distribution function

$$H(x) = \lambda\, F(x) + (1-\lambda)\, G(x) \quad .$$

We also use the sample quantile functions,

$$\tilde{Q}_X(u) = \tilde{F}^{-1}(u),\; \tilde{Q}_Y(u) = \tilde{G}^{-1}(u),\; \text{and } \tilde{Q}_H(u) = \tilde{H}^{-1}(u) \;, \tag{1.3}$$

where in general the quantile function Q is defined by

$$Q(u) = F^{-1}(u) = \inf_x \{x : F(x) \ge u\} \;,\; 0 \le u \le 1 \;. \tag{1.4}$$

We also define a sample comparison distribution function by

$$\tilde{D}(u) = \tilde{F}[\tilde{H}^{-1}(u)] \;,\; 0 \le u \le 1 \quad , \tag{1.5}$$

and the population comparison distribution function corresponding to $\tilde{D}(u)$ by

$$D(u) = F[H^{-1}(u)] \ , \quad 0 \leq u \leq 1 \quad . \tag{1.6}$$

Alternative definitions for a comparison function are FG^{-1}, GF^{-1}, $G^{-1}F$ and $F^{-1}G$. Doksum and Sievers (1976) and others have studied some of these alternative comparison functions.

Switzer (1976), Doksum and Sievers (1976), Wilk and Gnanadesikan (1968), Doksum (1974) and Steck, Zimmer, and Williams (1974) have also studied comparison functions. However, here $D(u)=F[H^{-1}(u)]$ is preferred because it tends to have more jump points than any of the other forms. It is, in a sense, "smoother" than any of the others. Furthermore, $\{R_{Ni};\ 1 \leq i \leq N\}$ the set of relative ranks of the X sample are given by

$$R_{Ni} = m\ \tilde{F}[\tilde{H}^{-1}\ (\tfrac{i}{N})] \quad .$$

as noted in Pyke and Shorack (1968). In sections 4.1-4.3 we will provide some data analytic comparisons of our approach with those using alternative comparison functions. These comparisons will also emphasize the theoretical differences.

In modelling D(u) and Q(u) we require the density-quantile function defined by

$$fQ(u) = f[Q(u)] = F'[Q(u)] \quad , \tag{1.7}$$

and the score function defined by

$$J(u) = -\ (fQ)'(u) \quad , \tag{1.8}$$

to exist for all results presented in this work.

We further define some models for comparing the two samples. Under H_o: F=G and alternatives close to the null hypothesis in location and scale we use Parzen's (1980) model for D(u) defined by

$$D(u)-u=(1-\lambda)\{\theta f_oQ_o(u)+\psi Q_o(u)f_oQ_o(u)\} \quad , \tag{1.9}$$

where $\theta = \frac{\mu_2-\mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$. We also compare the two samples by a model of $\Delta_Q(u) = Q_Y(u) - Q_X(u)$, the difference of the quantile functions. This model will be suggested by Theorem 1.1 as

$$f_oQ_o(u)\Delta_Q(u)=(\mu_2-\mu_1)\, f_oQ_o(u)+(\sigma_2-\sigma_1)Q_o(u)f_oQ_o(u) \; , \tag{1.10}$$

which is valid under all location and scale alternatives to H_o:F=G. We denote estimators of D(u) and $\Delta_Q(u)$ as $\hat{D}(u)$ and $\hat{\Delta}_Q(u)$, respectively.

Estimators of D(u) and $\Delta_Q(u)$ can be obtained using the results of continuous parameter time series regression estimation developed by Parzen (1961, 1967). A detailed discussion of these estimators is given in sections 2 and 6 and is essentially taken from Parzen (1979), section 9 and 10, and Parzen (1980).

We define a Brownian bridge or a tied down Weiner process to be a normal process denoted by

$$\{B(u) \;,\; 0 \leq u \leq 1\} \; , \tag{1.11}$$

which has zero mean and covariance kernel

$$K_B(u_1,u_2) = \min\,(u_1,u_2) - u_1u_2 \quad . \tag{1.12}$$

Finally, we define the reproducing kernel Hilbert space (RKHS) of B(u) to be the space of L_2 differentiable functions with inner product

$$\langle f,g\rangle_{p,q} = \int_p^q f'(u)g'(u)\,du + \frac{1}{p}f(p)g(p) + \frac{1}{1-q}f(q)g(q) . \qquad (1.13)$$

Throughout this work we denote weak convergence of a process by "$\xrightarrow{L}$" and convergence in distribution by "$\xrightarrow{D}$".

1.1.2 Questions to be Addressed

One desires to infer from the samples how the populations for the two samples differ. The distribution functions F(x) and G(x) each have, in general, an infinite number of parameters and it is our task to summarize or characterize the differences in these distribution functions. The quantile function has advantages in this regard as remarked in Parzen (1979) and Wilk and Gnanadesikan (1968).

One explanation of its statistical virtues is the fact that

$$\tilde{Q}(u) = X_{(j)} \ , \text{ for } \frac{j-1}{n} < u \leq \frac{j}{n} \quad ,$$

where $X_{(j)}$ is the j^{th} order statistic. The order statistics are the most universal set of sufficient statistics since all sufficient statistics are a function of the order statistics.

The problems we address are illustrated by the t-test. If one assumes the data are normally distributed with $\sigma_1 = \sigma_2$, one obtains an exact solution (t-test) to a well-posed problem. Of course, these assumptions are usually only approximately true. Thus, the t distribution, which gives both a test of $H_o: \mu_1 = \mu_2$ and a confidence interval for $\mu_1 - \mu_2$, provides exact solutions to approximate problems. If $\sigma_1 \neq \sigma_2$, we have the Behrens-Fisher problem, which is usually more realistic and currently has no exact solution.

The nonparametric problem considered in this dissertation assumes that the data are not known to be normally distributed. One problem is then to choose from a collection of F_o functions those which best fit the data. Another problem is to develop techniques to estimate and test hypotheses concerning the parameters $\theta = \frac{\mu_2-\mu_1}{\sigma_1}$, $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$, $\mu_2 - \mu_1$, and $\sigma_2 - \sigma_1$. These techniques then provide tests of H_o: F=G and an estimator, $\hat{D}(u)$, of D(u).

Through $\hat{D}(u) - \tilde{D}(u)$ we provide techniques to determine:

(1) whether the two samples differ in location and scale parameters for a given f_o, and

(2) whether the assumed density, f_o, can model the data well.

We implement and expand some of Parzen's (1979, 1980) results. In cases where several F_o may model the data, one can compare the various estimates of θ, ψ, $\mu_2 - \mu_1$, and $\sigma_2 - \sigma_1$. When quantitative or qualitative differences exist among the various F_o we will suggest larger samples for more power, or subject matter based selection of F_o rather than statistically based selection.

1.2 The Solution

We introduce the solution in this section and give the detailed implementation of the solution in sections 2.1 through 2.5 and 6.2.

The aim of the approach implemented in this work is to simultaneously estimate location and scale differences between two populations. Our approach emphasizes estimators of location and scale differences that are asymptotically optimal for the same underlying f_o. The approach may also provide diagnostics for skewness, long tails, bimodality, and estimates for nonconstant shifts in the various quantiles of a population using techniques from Parzen (1979). The approach begins with $D(u) = F[H^{-1}(u)]$ and its raw estimator $\tilde{D}(u) = \tilde{F}[\tilde{H}^{-1}(u)]$.

Since $F(x) = G(x)$ iff $H^{-1}(u) = F^{-1}(u) = G^{-1}(u)$ iff $D(u) = F[H^{-1}(u)] = u$, the comparison function $\tilde{D}(u)$ can be used to test H_o: $F(x)=G(x)$ by testing H_o: $D(u)=u$. The asymptotic distribution of $\tilde{D}(u)$ under H_o: $F=G$, is given by

$$\sqrt{N}\ [\tilde{D}(u)-u] \xrightarrow{L} \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u) = cB(u)$$

where $\lambda = \frac{m}{N} \to \lambda_o$, $0 < \lambda_o < 1$. A proof of this fact is outlined in Parzen (1980), and essentially given in Pyke and Shorack (1968).

Parzen's (1980) representation

$$\sqrt{N}[\tilde{D}(u)-u]=(1-\lambda)[\theta f_oQ_o(u)+\psi Q_o(u)f_oQ_o(u)] + cB(u)$$

is adopted here. In essence, this is the result of a linear Taylor series expansion for D(u) which we discuss in section 2 in greater detail.

As a result, with Parzen's (1961, 1967) results, we simultaneously estimate θ and ψ from a commonly assumed f_o as

$$(1-\lambda)\begin{bmatrix}\hat{\theta}\\ \hat{\psi}\end{bmatrix} = \begin{bmatrix}\langle f_oQ_o, f_oQ_o\rangle & \langle f_oQ_o, Q_o(f_oQ_o)\rangle \\ \langle f_oQ_o, Q_o(f_oQ_o)\rangle & \langle Q_o(f_oQ_o), Q_o(f_oQ_o)\rangle\end{bmatrix}^{-1}$$

$$\cdot \begin{bmatrix}\langle f_oQ_o, \tilde{D}(u) - u\rangle \\ \langle Q_o(f_oQ_o), \tilde{D}(u) - u\rangle\end{bmatrix} ,$$

where

$$\langle f_1, f_2\rangle_{p,q} = \int_p^q f_1'(u)\, f_2'(u)\, du + \frac{1}{p} f_1(p) f_2(p) + \frac{1}{1-q} f_1(q) f_2(q)$$

and

$$\langle f_1, f_2\rangle = \lim_{\substack{p\to 0\\ q\to 1}} \langle f_1, f_2\rangle_{p,q} .$$

Estimators obtained when one uses $\langle f_1, f_2\rangle_{p,q}$; $0 < p < q < 1$, are briefly mentioned in section 2.4 where we obtain trimmed or truncated estimators of θ and ψ for the exponential density. Also, note that the inner product $\langle f_1, f_2\rangle_{p,q}$ exists for many more density functions than does $\langle f_1, f_2\rangle$ since the latter requires that, for $j=1,2$, $f_j(p) \to 0$ as $p \to 0, 1$. The computational formulas for $\hat{\theta}$ and $\hat{\psi}$ are surprisingly simple for many densities. The similarities of tests based on $\hat{\theta}$ and $\hat{\psi}$ to other tests will be established in sections 5.1 through 5.4. Table 1 gives $\hat{\theta}$ and $\hat{\psi}$ for several f_o densities. Before proceeding to the derivation of these estimators we will consider a model for the quantile functions.

1a. Relationship of Classical Nonparametric Test Statistics to $\hat{\theta}$

	Name	Test Statistic	Estimator	f_o
Location:	Van der Waerden	$\sum_{i=1}^{m} \Phi^{-1}\left(\frac{R_i}{N+1}\right)$	$= -m\hat{\theta}$	normal
	Wilcoxon	$\sum_{i=1}^{m} R_i$	$= 2m - \frac{m(N+1)}{6}\hat{\theta}$	logistic
	Median (Mood)	$\sum_{i=1}^{m} \mathrm{sign}[R_i - \frac{1}{2}(N+1)] =$	$m\hat{\theta}$	double exponential
	none	none	$\hat{\theta} = +2/m \sum_{i=1}^{m} \sin\left(2\pi \frac{R_i}{N+1}\right)$	Cauchy
	none "Ansari-Bradley"	none	$\hat{\theta} = \frac{3}{m} \sum_{i=1}^{m} \mathrm{sign}\left(\frac{1}{2} - \frac{R_i}{N+1}\right) \min\left(\frac{R_i}{N+1}, 1 - \frac{R_i}{N+1}\right)$	$f_o(x) = \frac{1}{2}(1+\lvert x \rvert)^{-2}$
	none (Quantile density)	none	$\hat{\theta} = \frac{3}{m}\left[\sum_{R_i \le \frac{1}{4}(N+1)} \left(\frac{R_i}{N+1}\right) + \sum_{R_i \ge \frac{3}{4}(N+1)} \left(\frac{R_i}{N+1} - 1\right)\right]$	$f_o(x) = 1, \ \lvert x \rvert \le \frac{1}{4}$; $= \frac{1}{16x^2}, \ \lvert x \rvert > \frac{1}{4}$

Note: R_i is the rank of X_i in the combined sample.

1b. Relationship of Classical Nonparametric Test Statistics to $\hat{\psi}$

	Name	Test	Estimator	f_o
Scale:	Klotz	$\sum_{i=1}^{m} [\Phi^{-1}(\frac{R_i}{N+1})]^2$	$= -2m(\hat{\psi} - \frac{1}{2})$	normal
	none (Wilcoxon density)	none	$\hat{\psi} = \frac{9}{m(3+\pi^2)} \sum_{i=1}^{m} \log(\frac{R_i/(N+1)}{1-R_i/(N+1)})[2(\frac{R_i}{N+1})-1]$	logistic
	none (Mood or Median density)	none	$\hat{\psi} = 1+\frac{1}{m} \sum_{i=1}^{m} \log 2[\min(\frac{R_i}{N+1}, 1-\frac{R_i}{N+1})]$	double exponential
	none	none	$\hat{\psi} = \frac{2}{5} - \frac{2}{5m} \sum_{i=1}^{m} \{-\sin[2\pi(\frac{R_i}{N+1})]\}\tan[\pi(\frac{R_i}{N+1} - \frac{1}{2})]$	Cauchy
	Ansari-Bradley or Siegel-Tukey	$\sum_{i=1}^{m} [\frac{1}{2}(N+1) - \left\lvert R_i - \frac{1}{2}(N+1) \right\rvert]$	$= \frac{m}{2}(N+1) - \frac{m(N+1)}{12}(\hat{\psi}-3)$	$f_o(x) = \frac{1}{2}(1+\lvert x \rvert)^{-2}$
	Quartile	$T =$ #of X obs. $\in (\tilde{H}^{-1}(.25), \tilde{H}^{-1}(.75))$	$= m(\frac{1}{2} - \hat{\psi})$	$f_o(x) = 1, \lvert x \rvert \le \frac{1}{4}$, $= \frac{1}{16x^2}, \lvert x \rvert \ge \frac{1}{4}$.

The quantile functions $\tilde{Q}_X$ and $\tilde{Q}_Y$ can also be plotted to compare the two samples. These plots and their box plots, as defined in Parzen (1979), give initial indications of skewness, bimodality, and differences in location and scale. A model for the differences at each quantile is given in the following theorem.

Theorem 1.1: If $\{X_i \,;\, i = 1, \ldots, m\}$ is a random sample from $F(x) = F_o(\frac{x-\mu_1}{\sigma_1})$ and $\{Y_j \,;\, j = 1, \ldots, n\}$ is an independent random sample from $G(x) = F_o(\frac{x-\mu_2}{\sigma_2})$, where f_o' exists, $f_o>0$ is continuous and tail monotone [see Parzen (1979), p. 116], then, as $N \to \infty$, such that $\lambda_N = \frac{m}{N} \to \lambda_o$ $(0 < \lambda_o < 1)$,

$$\sqrt{N}\, f_oQ_o(u)[\tilde{Q}_Y(u)-\tilde{Q}_X(u)-(\mu_2-\mu_1)-(\sigma_2-\sigma_1)Q_o(u)] \xrightarrow{L} c\, B(u),$$

where $c^2 = \lambda_o\sigma_1^2 + (1-\lambda_o)\,\sigma_2^2$ and $B(u)$ is a Brownian bridge.

Proof: From Parzen (1979), since f_o is tail monotone, we have

$$\sqrt{N}(f_oQ_o(u))(\tilde{Q}_Y(u)-\mu_2-\sigma_2Q_o(u)) \xrightarrow{L} (1-\lambda_o)^{\frac{1}{2}}\sigma_2\, B_2(u) \quad ,$$

as $N \to \infty$, and

$$-\sqrt{N}(f_oQ_o(u))(\tilde{Q}_X(u)-\mu_1-\sigma_1Q_o(u)) \xrightarrow{L} -\lambda_o^{\frac{1}{2}}\sigma_1\, B_1(u) \quad ,$$

as $N \to \infty$ where $B_1(u)$ and $B_2(u)$ are independent Brownian bridges. Thus, the independence of $\tilde{Q}_X$ and $\tilde{Q}_Y$ yields

$$\sqrt{N}(f_oQ_o(u))[(\tilde{Q}_Y(u)-\tilde{Q}_X(u)-(\mu_2-\mu_1)-(\sigma_2-\sigma_1)Q_o(u)] \xrightarrow{L} Z(u) \quad ,$$

where

$$Z(u) = (1-\lambda_o)^{\frac{1}{2}} \sigma_2 B_2(u) - \lambda_o^{\frac{1}{2}} \sigma_1 B_1(u) \quad .$$

Now, $E[Z(u)] = 0$, since $B_1(u)$ and $B_2(u)$ are zero mean normal processes. For $0 \leq u_1 \leq 1$ and $0 \leq u_2 \leq 1$, we have

$$\begin{aligned}
\mathrm{cov}\,[Z(u_1), Z(u_2)] &= \mathrm{cov}\,[(1-\lambda_o)^{\frac{1}{2}}\sigma_2 B_2(u_1)-\lambda_o^{\frac{1}{2}}\sigma_1 B_1(u_1) \quad , \\
&\qquad (1-\lambda_o)^{\frac{1}{2}}\sigma_2 B_2(u_2)-\lambda_o^{\frac{1}{2}}\sigma_1 B_1(u_2)] \\
&= E[(1-\lambda_o)\sigma_2^2 B_2(u_1)B_2(u_2)-\lambda_o^{\frac{1}{2}}\sigma_1(1-\lambda_o)^{\frac{1}{2}} \\
&\qquad \cdot\sigma_2 B_2(u_1)B_1(u_2)-\lambda_o^{\frac{1}{2}}\sigma_1(1-\lambda_o)^{\frac{1}{2}}\sigma_2 B_1(u_1) \\
&\qquad \cdot B_2(u_2)+\lambda_o\sigma_1^2 B_1(u_1)B_1(u_2)] \\
&= E[(1-\lambda_o)\sigma_2^2 B_2(u_1)B_2(u_2)]+ \\
&\qquad E[\lambda_o\sigma_1^2 B_1(u_1)B_1(u_2)] \\
&= (1-\lambda_o)\sigma_2^2 \,\mathrm{cov}[B_2(u_1),B_2(u_2)] + \\
&\qquad \lambda_o\sigma_1^2 \,\mathrm{cov}[B_1(u_1),B_1(u_2)] \\
&= [\lambda_o\sigma_1^2+(1-\lambda_o)\sigma_2^2][\min(u_1,u_2)-u_1u_2] \quad .
\end{aligned}$$

Since linear combinations of independent Gaussian process are Gaussian, we have

$$Z(u) = [\lambda_o\sigma_1^2 + (1-\lambda_o)\sigma_2^2]^{\frac{1}{2}}B(u) \quad .$$

We thus have a model for $\Delta_Q = Q_Y - Q_X$. We also emphasize that $\Delta_Q(u)$ seems to be a very interesting and interpretable function since it quantifies the differences between X and Y at every quantile, $u\epsilon(0,1)$. We will then be able to obtain diagnostics for how the different quantiles of the populations are changed by the treatment. Further, some commonly used diagnostics which are functionals of $\tilde{\Delta}_Q = \tilde{Q}_Y - \tilde{Q}_X$ are $\bar{X}_2 - \bar{X}_1 = \int_o^1 \tilde{\Delta}_Q(u)du$ and the difference in medians $\tilde{\Delta}_Q(\frac{1}{2})$. Furthermore, if $\Delta_Q \equiv k$, a constant, then $\sigma_1 = \sigma_2$ and $\mu_1 \neq \mu_2$ ($k = 0$ implies $\mu_1 = \mu_2$).

This approach will thus provide:

(1) tests of $\mu_1 = \mu_2$ for several f_o ,

(2) tests of $\sigma_1 = \sigma_2$ for several f_o ,

(3) simultaneous tests for (1) and (2) for several common f_o for each sample,

(4) estimators for the parameters of (1), (2), and (3),

(5) models for estimating the difference at all quantiles, $\Delta_Q(u)$,

(6) graphical comparisons of the two samples ,

(7) a basis for theory on similar results for skewed and bimodal data ,

(8) a basis for theory on similar results using trimmed estimators, i.e., inner products with $0 < p \leq u \leq q < 1$.

1.3 Contributions of this Research

We implement and extend the techniques of Parzen (1979, 1980) by:

(1) giving calculation formulas for $\hat{\theta}$ and $\hat{\psi}$ for seven underlying densities;

(2) using the relation of $\hat{\theta}$ and $\hat{\psi}$ to other linear rank test statistics to provide finite sample size tests and parameter estimates based on $\hat{\theta}$ and $\hat{\psi}$;

(3) providing calculation formulas for the two parameter exponential distribution for truncated estimates of θ and ψ in the two sample problem;

(4) proving the asymptotic normality of $\underline{\hat{D}}(\underline{u}) = [\hat{D}(u_1), \hat{D}(u_2), \ldots, \hat{D}(u_k)]'$ for fixed $\{u_i, i = 1, \ldots, k\}$;

(5) proving the asymptotic normality of $\underline{\hat{D}}(\underline{u}) - \underline{\tilde{D}}(\underline{u})$ for fixed $\{u_i, i=1, \ldots, k\}$, which we use to select an underlying f_o ;

(6) deriving a model for $\Delta_Q(u) = Q_Y(u) - Q_X(u)$, the differences at each quantile, u ;

(7) giving estimation formulas for $\mu_2 - \mu_1$ and $\sigma_2 - \sigma_1$ simultaneously based on $\tilde{Q}_Y$ and $\tilde{Q}_X$;

(8) finding the asymptotic distribution for $\hat{\Delta}_Q(u) = \hat{Q}_Y(u) - \hat{Q}_X(u)$ at fixed u ; and

(9) providing graphical comparison techniques via $\hat{D}(u) - \tilde{D}(u)$ and $\hat{\Delta}_Q(u) - \tilde{\Delta}_Q(u)$.

We emphasize that all the theoretical contributions in this work are based on a location scale difference of two independent random samples with an assumed underlying f_o family common to both populations.

2. STATISTICAL INFERENCE BASED ON $\hat{D}(u)$, $\hat{\theta}$, and $\hat{\psi}$

In this section we present the basic method for making inferences about two populations given random samples from each which are independent. We assume the two populations and samples are as given by definitions in section 1.1.1. In section 2.1 we outline the results of Parzen (1961, 1967) which provide the theory to suggest the estimators $\hat{\theta}$, $\hat{\psi}$, and $\hat{D}(u)$ for $\theta = \frac{\mu_2-\mu_1}{\sigma_1}$, $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$, and $D(u) = F[H^{-1}(u)]$ also defined in section 1.1.1.

In section 2.2 we use these results to obtain the computational formulas for $\hat{\theta}$, $\hat{\psi}$, and $\hat{D}(u)$ which lead to the relationships given earlier (Table 1a/b, p. 9) for several density functions f_o. In section 2.3 some large sample distribution theory for the estimators obtained in section 2.2 is discussed.

Since the methods of section 2.1 through 2.3 do not apply for all choices of f_o, in section 2.4 we show how we may use truncated estimators using formula (1.13) for the particular case of the two parameter exponential f_o. Finally, in section 2.5 we describe some finite sample size distributional results for the estimators of sections 2.1-2.4.

2.1 Time Series Regression and Preliminaries

Using the definitions in section 1.1.1, Parzen (1980) has suggested a model for D(u) which we use to obtain estimators of

$\theta = \frac{\mu_2-\mu_1}{\sigma_1}$, $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$ and D(u). The model is particularly suggested when θ and ψ are near zero, i.e., the remainder terms of a Taylor series expansion are small. For $D(u) = FH^{-1}(u)$ when θ and ψ are small, Parzen (1980) suggests that we may use

$$\tilde{D}(u)-u = (1-\lambda)[\theta \; f_oQ_o(u)+\psi \, Q_o(u) \; f_oQ_o(u)]+(\frac{1-\lambda}{N\lambda})^{\frac{1}{2}} \; B(u) \; .$$

This model briefly is a Taylor series expansion of F_o about $(\frac{x-\mu_1}{\sigma_1})$ when θ and ψ are small. A sketch of Parzen's justification for

$$D(u)-u = (1-\lambda)[\theta \; f_oQ_o(u)+\psi \, Q_o(u) \; f_oQ_o(u)]$$

is given below.

<u>Derivation of Parzen's Representation for D(u)-u</u> .

Since $\theta\sigma_1 = \mu_2-\mu_1$ and $\sigma_1 \; (1 + \psi) = \sigma_2$, we have

$$G(x) = F_o(\frac{x-\mu_2}{\sigma_2}) = F_o(\frac{x-\mu_1-\theta\sigma_1}{\sigma_1(1+\psi)}) \quad .$$

Since $(1+\psi)^{-1} \doteq 1-\psi$ (when ψ^2 is small) and $\theta\psi \doteq 0$, we have

$$G(x) \doteq F_o[(\frac{x-\mu_1}{\sigma_1})(1-\psi) - \theta]$$

$$= F_o\{(\frac{x-\mu_1}{\sigma_1}) - [\theta + \psi(\frac{x-\mu_1}{\sigma_1})]\} \quad .$$

A linear Taylor series expansion of this representation of G(x) about $(\frac{x-\mu_1}{\sigma_1})$ gives

$$G(x) \doteq F_o\left(\frac{x-\mu_1}{\sigma_1}\right) + f_o\left(\frac{x-\mu_1}{\sigma_1}\right)\left[-\theta - \psi\left(\frac{x-\mu_1}{\sigma_1}\right)\right] .$$

Substituting this in $H(x) = \lambda F(x) + (1-\lambda)\, G(x)$ gives

$$H(x) \doteq \lambda F(x)+(1-\lambda)F(x)-(1-\lambda)f_o\left(\frac{x-\mu_1}{\sigma_1}\right)\left[\theta+\psi\left(\frac{x-\mu_1}{\sigma_1}\right)\right]$$

$$= F(x)-(1-\lambda)f_o\left(\frac{x-\mu_1}{\sigma_1}\right)\left[\theta+\psi\left(\frac{x-\mu_1}{\sigma_1}\right)\right] .$$

Letting $x = H^{-1}(u)$ and rearranging terms, yields

$$D(u) = u + (1-\lambda)f_o\left(\frac{H^{-1}(u)-\mu_1}{\sigma_1}\right)\left[\theta+\psi\left(\frac{H^{-1}(u)-\mu_1}{\sigma_1}\right)\right] .$$

Since $[f_o(\frac{H^{-1}(u)-\mu_1}{\sigma_1})-f_o(\frac{F^{-1}(u)-\mu_1}{\sigma_1})]\theta$ and $[F^{-1}(u)-H^{-1}(u)]\,\psi$ are an order smaller than θ and ψ, as θ and ψ go to zero we have

$$D(u)-u \doteq (1-\lambda)f_o\left(\frac{F^{-1}(u)-\mu_1}{\sigma_1}\right)\left[\theta+\psi\left(\frac{F^{-1}(u)-\mu_1}{\sigma_1}\right)\right]$$

$$= (1-\lambda)f_oQ_o(u)\left[\theta+\psi\, Q_o(u)\right]$$

The error term of $(\frac{1-\lambda}{\lambda})^{\frac{1}{2}}\, B(u)$ that Parzen suggests is adopted from Theorem 4.1 of Pyke and Shorack (1968) with the constants of their Lemma 3.1 and equation 3.7. We give the result we need in Theorem 2.1.

<u>Theorem 2.1</u>: If the conditions of Theorem 4.1 in Pyke and Shorack (1968) hold and $F(x) \approx G(x)$ for F and G as defined in section 1.1.1, then

$$\sqrt{N}\ [\tilde{D}(u)-u] \xrightarrow{L} \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u) \ ,$$

where $\lambda_N = \frac{m}{N} \to \lambda_o (0 < \lambda_o < 1)$ as $N \to \infty$ and $B(u)$ is a Brownian bridge.

Proof: Pyke and Shorack (1968) define

$$L_N(u) = \sqrt{N}\ [\tilde{D}(u)-D(u)] = \sqrt{N}\ \{\tilde{F}[\tilde{H}^{-1}(u)] - F[H^{-1}(u)]\} \ ,$$

and show for

$$L_N{}'\ (u) = L_n(u) \ , \ \frac{1}{N} \le u \le 1 \ ,$$

$$= 0 \ , \ 0 \le u < \frac{1}{N} \ ,$$

that $\rho(L_N, L_N') \xrightarrow{a.s.} 0$, where ρ is the uniform metric, and

$$L_N{}'\ (u) \xrightarrow{L} L_o(u) \ .$$

Under H_o: $F = G$, we have

$$L_o(u) = (1-\lambda_o)\ \{\lambda_o^{-\frac{1}{2}} B_1(u) - (1-\lambda_o)^{-\frac{1}{2}} B_2(u)\} \ ,$$

where $B_1(u)$ and $B_2(u)$ are independent Brownian bridges. As in our Theorem 1.1, we have

$$L_o(u) = (1 - \lambda_o)\ \{c\ B(u)\} \ ,$$

where $c^2 = \lambda_o^{-1} + (1-\lambda_o)^{-1} = [\lambda_o(1-\lambda_o)]^{-1}$.

Therefore,

$$L_o(u) = \frac{(1-\lambda_o)}{\lambda_o^{\frac{1}{2}}(1-\lambda_o)^{\frac{1}{2}}}\ B(u) = \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u) \ ,$$

and

$$\sqrt{N}\,[\tilde{D}(u) - u] \xrightarrow{L} \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u) \quad .$$

We note that although this error term is only shown to be correct when $\theta = \psi = 0$, we shall assume that it is approximately correct for θ and ψ close to zero.

That is, we use the Parzen (1961, 1967) results to obtain estimates of θ and ψ under H_o: $\theta = \psi = 0$ which we assume will be useful for θ and ψ near 0 also. Exactly under what conditions this is justified is an open research problem. One can calculate estimators for all continuous f_oQ_o and $Q_o(f_oQ_o)$ in the RKHS of B(u) (see section 1) with

$$\langle f_1, f_2 \rangle = \lim_{\substack{p \to 0 \\ q \to 1}} \langle f_1, f_2 \rangle_{f,q} \quad ,$$

where

$$\langle f_1, f_2 \rangle_{p,q} = \int_p^q f_1'(u) f_2'(u)\,du + \frac{1}{p} f_1(p) f_2(p) + \frac{1}{1-q} f_1(q) f_2(q) \quad .$$

The conditions of Lemma 2.1 are sufficient for the estimation of θ and ψ using Parzen (1961, 1967). This gives

$$(1 - \lambda) \binom{\hat{\theta}}{\hat{\psi}} = \Sigma^{-1} \underline{g}$$

where

$$\Sigma = \begin{bmatrix} \langle f_oQ_o, f_oQ_o \rangle & \langle f_oQ_o, Q_o(f_oQ_o) \rangle \\ \langle Q_o(f_oQ_o), f_oQ_o \rangle & \langle Q_o(f_oQ_o), Q_o(f_oQ_o) \rangle \end{bmatrix}$$

and

$$\underline{g} = \begin{bmatrix} \langle f_oQ_o, \tilde{D}(u)-u \rangle \\ \\ \langle Q_o(f_oQ_o), \tilde{D}(u)-u \rangle \end{bmatrix}$$

as in section 1.2, for a solution to the normal equations

$$(1 - \lambda)\ \Sigma\binom{\hat{\theta}}{\hat{\psi}} = \underline{g} \quad .$$

The estimators $\hat{\theta}$ and $\hat{\psi}$ then give

$$\hat{D}(u) = u + (1-\lambda)[\hat{\theta}\ f_oQ_o(u) + \hat{\psi}\ Q_o(u)f_oQ_o(u)] \ ,$$

using Parzen's model for D(u).

2.2 Calculation of $\hat{\theta}$, $\hat{\psi}$, and $\hat{D}(u)$

In this section we give some lemmas useful in calculating Σ and $\underline{g}$ for various f_o. We then calculate $\hat{\theta}$ and $\hat{\psi}$ for seven different f_o densities.

We consider here the following $f_o(x)$; $(-\infty < x < \infty$, unless otherwise specified)

Normal $\quad f_o(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ , \quad F_o(x) = \Phi(x) \quad .$

Logistic $\quad f_o(x) = e^x(1 + e^x)^{-2} \ , \quad F_o(x) = (1 + e^x)^{-1} \quad .$

Cauchy $\quad f_o(x) = \pi^{-1}(1+x^2)^{-1}$, $\quad F_o(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)$.

Double Exponential $\quad f_o(x) = \frac{1}{2}e^{x},\ x \le 0$, $\quad F_o(x) = \frac{1}{2}e^{x},\ x \le 0$,
$\quad = \frac{1}{2}e^{-x},\ x \ge 0$, $\quad = 1 - \frac{1}{2}e^{-x},\ x \ge 0$.

"Ansari-Bradley" $\quad f_o(x) = \frac{1}{2}(1+|x|)^{-2}$, $\quad F_o(x) = \frac{1}{2}(1-x)^{-1},\ x \le 0$,
$\quad = \frac{1}{2} + \frac{1}{2}[1-(1+x)^{-1}],\ x > 0$.

"Quartile" $\quad f_o(x) = 1\ ,\ |x| \le \frac{1}{4}$, $\quad F_o(x) = -\frac{1}{16x},\ x \le -\frac{1}{4}$,
$\quad = \frac{1}{16x^2},\ |x| > \frac{1}{4}$, $\quad = \frac{1}{2} + x,\ x \in (-\frac{1}{4}, \frac{1}{4})$,
$\quad = 1 - \frac{1}{16x},\ x > \frac{1}{4}$.

Exponential $\quad f_o(x) = e^{-x},\ x > 0$, $\quad F_o(x) = 1 - e^{-x},\ x \ge 0$,
$\quad = 0,\ x < 0$, $\quad = 0,\ x < 0$.

The formulas for Σ and $\underline{g}$ require several inner products. The calculation of these inner products can be simplified for many f_o by using the following lemmas.

Lemma 2.1: If f_oQ_o and $Q_o(f_oQ_o)$ are L_2 differentiable functions, then

$$\text{(1)} \quad \langle f_oQ_o, f_oQ_o \rangle = \int_0^1 J_{o\cdot}^2(u)du$$

when

$$\text{(i)} \quad 0 = \lim_{p\to 0} \frac{[f_oQ_o(p)]^2}{p} = \lim_{p\to 0} \frac{[f_oQ_o(1-p)]^2}{p} ,$$

(2) $\quad \langle Q_o(f_oQ_o),\ Q_o(f_oQ_o)\rangle = \int_0^1 [1 - Q_o(u)J_o(u)]^2\,du$

when

(ii) $\quad 0 = \lim_{p\to 0} \dfrac{[Q_o(p)f_oQ_o(p)]^2}{p} = \lim_{p\to 0} \dfrac{[Q_o(1-p)f_oQ_o(1-p)]^2}{p}$,

(3) $\quad \langle f_oQ_o,\ Q_o(f_oQ_o)\rangle = \int_0^1 Q_o(u)J_o^{\,2}(u)du - \int_0^1 J_o(u)du$

when (i) and (ii),

(4) $\quad \langle f_oQ_o,\ \tilde{D}(u)-u\rangle = \int_0^1 J_o(u)du - \frac{1}{m}\sum_{i=1}^{m} J_o(\frac{R_i}{N+1})$

when (i), and

(5) $\quad \langle Q_o(f_oQ_o),\tilde{D}(u)-u\rangle = \int_0^1 Q_o(u)J_o(u)du - \frac{1}{m}\sum_{i=1}^{m} Q_o(\frac{R_i}{N+1})J_o(\frac{R_i}{N+1})$

when (ii).

Proof: [Adapted from Parzen (1979, 1980)]

(1) By definition, $J_o(u) = -(f_oQ_o)'(u)$ which gives

$$\langle f_oQ_o,f_oQ_o\rangle = \lim_{\substack{p\to 0\\ q\to 1}} \langle f_oQ_o,f_oQ_o\rangle_{p,q}$$

$$= \lim_{\substack{p\to 0\\ q\to 1}}\left[\int_p^q [-J_o(u)]^2du + \frac{[f_oQ_o(p)]^2}{p} + \frac{[f_oQ_o(q)]^2}{1-q}\right]$$

$$= \int_0^1 J_o^{\,2}(u)du$$

when (i).

(2) Since $q_o(u) = 1/f_oQ_o(u)$, and

$$[Q_o(u)f_oQ_o(u)]' = Q_o(u)(f_oQ_o)'(u) + f_oQ_o(u)Q_o'(u)$$

$$= -Q_o(u)J_o(u) + f_oQ_o(u)q_o(u)$$

$$= 1 - Q_o(u)J_o(u) ,$$

$$<Q_o(f_oQ_o),Q_o(f_oQ_o)> = \int_0^1 [1-Q_o(u)J_o(u)]^2 \, du$$

when (ii).

(3) Similarly,

$$<f_oQ_o,Q_o(f_oQ_o)> = \int_0^1 [-J_o(u)][1-Q_o(u)J_o(u)]du$$

$$= \int_0^1 Q_o(u)J_o^{\,2}(u)du - \int_0^1 J_o(u)du$$

when (i) and (ii).

(4) Similarly, since $\tilde{D}(1) = 1$ and $\tilde{D}(0) = 0$,

$$<f_oQ_o,\tilde{D}(u)-u> = \int_0^1 [-J_o(u)]d[\tilde{D}(u)-u]$$

$$= \int_0^1 J_o(u)du - \int_0^1 J_o(u)d\,\tilde{D}(u)$$

$$= \int_0^1 J_o(u)du - \frac{1}{m}\sum_{i=1}^{m} J_o\left(\frac{R_i}{N+1}\right) ,$$

recalling that $\tilde{D}(u)$ has jumps $\frac{1}{m}$ at $u = \frac{R_i}{N+1}$.

(5) Also,

$$\langle Q_o(f_oQ_o),\tilde{D}(u)-u\rangle = \int_0^1 [1-Q_o(u)J_o(u)]d[\tilde{D}(u)-u]$$

$$= \int_0^1 d\tilde{D}(u)-\int_0^1 Q_o(u)J_o(u)d\tilde{D}(u)-1+\int_0^1 Q_o(u)J_o(u)du$$

$$= \int_0^1 Q_o(u)J_o(u)du - \frac{1}{m}\sum_{i=1}^{m} Q_o(\frac{R_i}{N+1})J_o(\frac{R_i}{N+1}) \quad .$$

Remark: The tail conditions of Lemma 2.1 essentially are the conditions needed besides L_2 differentiability for f_oQ_o and $Q_o(f_oQ_o)$ to be in the RKHS of $B(u)$ for $p=1-q=0$. To make this clear is why we include Lemma 2.1. This lemma is used to show $\hat{\theta}$ and $\hat{\psi}$ are linear rank statistics.

Lemma 2.2: If in addition to f_oQ_o and $Q_o(f_oQ_o)$ being L_2 differentiable, we have that f_o is symmetric, then

(1) $\langle f_oQ_o,Q_o(f_oQ_o)\rangle = 0$

when (i) and (ii) of Lemma 2.1 and

(2) $\langle f_oQ_o,\tilde{D}(u)-u\rangle = -\frac{1}{m}\sum_{i=1}^{m} J_o(\frac{R_i}{N+1})$

when (i) of Lemma 2.1

Proof: Since f_o symmetric is equivalent to $f_oQ_o(1-u) = f_oQ_o(u)$ or $J_o(1-u) = -J_o(u)$ or $Q_o(1-u) = Q_o(u)$ and Lemma 2.1 holds, we have

(1) $$\langle f_oQ_o,Q_o(f_oQ_o)\rangle = \int_0^1 Q_o(u)J_o^{\,2}(u)du-\int_0^1 J_o(u)du$$

$$= \int_0^{\frac{1}{2}} Q_o(u)J_o^{\,2}(u)du+\int_0^{\frac{1}{2}} J_o(u)du$$

$$-\int_0^{\frac{1}{2}} Q_o(u)J_o^{\,2}(u)du-\int_0^{\frac{1}{2}} J_o(u)du$$

$$= 0 \; ,$$

and

$$(2)\ \langle f_oQ_o, \tilde{D}(u)-u\rangle = -\frac{1}{m}\sum_{i=1}^{m} J_o\left(\frac{R_i}{N+1}\right), \text{ since } \int_0^1 J_o(u)du = 0.$$

Theorem 2.2: If f_oQ_o and $Q_o(f_oQ_o)$ are in the RKHS of B(u) with $p = 1 - q = 0$, $F = G$, and f_o is symmetric with the tail conditions of Lemma 2.1, then

$$(1-\lambda)\hat{\theta} = [\int_0^1 J_o^2(u)du]^{-1}[-\frac{1}{m}\sum_{i=1}^{m} J_o(\frac{R_i}{N+1})]$$

and

$$(1-\lambda)\hat{\psi} = \{\int_0^1 [1-Q_o(u)J_o(u)]^2 du\}^{-1}[\int_0^1 Q_o(u)J_o(u)du - \frac{1}{m}\sum_{i=1}^{m} Q_o(\frac{R_i}{N+1})J_o(\frac{R_i}{N+1})\ ]$$

Proof: Since Lemma 2.1 and 2.2 give the terms of Σ and $\underline{g}$, we have $(1-\lambda)\binom{\hat{\theta}}{\hat{\psi}} = \Sigma^{-1}\ \underline{g}$ as given in Theorem 2.2.

Note that our Σ for the two sample case is the same as the one sample Σ in Parzen (1979) and Eubank (1979). In order to carry out the estimation of θ and ψ as given in the above theorem, we need $f_oQ_o(u)$, $Q_o(u)$, and $J_o(u)$ for each density. They are given in Table 2. We obtain the results in Table 2 for the normal, logistic, and Cauchy densities as in Parzen (1979) and Eubank (1979). The others are also obtained by using

$$Q_o(u) = F_o^{-1}(u), (f_oQ_o)(u) = f_o[F_o^{-1}(u)] \text{ or } f_oQ_o(u) = 1/Q_o'(u)$$

$$\text{or } J_o(u) = -(f_oQ_o)'(u) \quad .$$

2. Density-Quantile, Quantile and Score Functions

f_0	f_0Q_0	Q_0	J_0
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\|\Phi^{-1}(u)\|^2}$	$\Phi^{-1}(u)$	$\Phi^{-1}(u)$
Logistic	$u(1-u)$	$\log \frac{u}{1-u}$	$2u-1$
Cauchy	$\frac{1}{\pi} \sin^2(\pi u)$	$\tan[\pi(u-\frac{1}{2})]$	$-\sin(2\pi u)$
Double Exponential	$u \;, u \leq \frac{1}{2}$ $1-u, u \geq \frac{1}{2}$	$\log 2u \;, u \leq \frac{1}{2},$ $-\log 2(1-u), u \geq \frac{1}{2}$	$-1, u < \frac{1}{2},$ $1, u > \frac{1}{2}$
Ansari-Bradley density	$2u^2 \;, u \leq \frac{1}{2}$ $2(1-u)^2, u > \frac{1}{2}$	$1-\frac{1}{2}u^{-1}, u \leq \frac{1}{2}$ $-1+\frac{1}{2}(1-u)^{-1}, u > \frac{1}{2}$	$-4u, u \leq \frac{1}{2},$ $4(1-u), u > \frac{1}{2}$
Quartile density	$16u^2, \; u < \frac{1}{4},$ $0, \; u \in (\frac{1}{4},\frac{3}{4}),$ $16(1-u)^2, u > \frac{3}{4}$	$-\frac{1}{16}u^{-1}, u < \frac{1}{4}$ $u-\frac{1}{2}, u \in (\frac{1}{4},\frac{3}{4}),$ $\frac{1}{16}(1-u)^{-1}, u \geq \frac{3}{4}$	$-32u, u < \frac{1}{4},$ $0, \; u \in (\frac{1}{4},\frac{3}{4}),$ $32(1-u), u > \frac{3}{4}$
Exponential	$1 - u$ *	$\log (1-u)^{-1}$	1

*Not in RKHS of Brownian Bridge process for $p = 1 - q = 0$.

Thus, we obtain $(1-\lambda)\binom{\hat{\theta}}{\hat{\psi}} = \Sigma^{-1}\,\underline{g}$ as above.

Theorem 2.3: The estimators of θ and ψ in

$$\tilde{D}(u)-u=(1-\lambda)[\theta f_oQ_o(u)+\psi Q_o(u)f_oQ_o(u)]+\left(\frac{1-\lambda_o}{N\lambda_o}\right)^{\frac{1}{2}} B(u)$$

are given in Table 3 for the seven densities.

Proof: Eubank (1979) has given Σ for the normal, logistic and Cauchy f_o. Since the tail conditions hold, we have

(Normal)

$$<f_oQ_o,f_oQ_o> = \int_0^1 J_o^2(u)du=\int_0^1 [\Phi^{-1}(u)]^2du=\int_{-\infty}^{\infty}x^2f(x)dx=1$$

and

$$<f_oQ_o,Q_o(f_oQ_o)> = 0 \text{ since } f_o \text{ is symmetric}$$

$$<Q_o(f_oQ_o),Q_o(f_oQ_o)> = \int_0^1 \{1-[\Phi^{-1}(u)]^2\}du$$

$$= 1-2\int x^2f(x)dx + \int x^4f(x)dx = 2.$$

Then,

$$\Sigma^{-1} = \begin{bmatrix}1 & 0\\ 0 & \frac{1}{2}\end{bmatrix} .$$

For $\underline{g}$ we have

$$<f_oQ_o,\tilde{D}(u)-u> = -\frac{1}{m}\sum_{i=1}^{m}\Phi^{-1}\left(\frac{R_i}{N+1}\right)$$

and

$$<Q_o(f_oQ_o),\tilde{D}(u)-u> = 1-\frac{1}{m}\sum_{i=1}^{m}\left[\Phi^{-1}\left(\frac{R_i}{N+1}\right)\right]^2 ,$$

3. Computational Formulas for $\hat{\theta}$ and $\hat{\psi}$

Density	$(1-\lambda)\ \hat{\theta}$	$(1-\lambda)\ \hat{\psi}$
Normal	$-\frac{1}{m}\Sigma\ \Phi^{-1}(\frac{R_i}{N+1})$	$\frac{1}{2}-\frac{1}{2m}\ \Sigma[\Phi^{-1}(\frac{R_i}{N+1})]^2$
Logistic	$3-\frac{6}{m}\ (\frac{R_i}{N+1})$	$-(\frac{9}{3+\pi^2})\frac{1}{m}\Sigma\log(\frac{\frac{R_i}{N+1}}{1-\frac{R_i}{N+1}})[2(\frac{R_i}{N+1})-1]$
Cauchy	$\frac{2}{m}\ \Sigma\ \sin(2\pi\ \frac{R_i}{N+1})$	$\frac{2}{5}-\frac{2}{5m}\ \Sigma\sin[2\pi(\frac{R_i}{N+1})]\tan[\pi(\frac{R_i}{N+1}-\frac{1}{2})]$
Double Exponential	$\frac{1}{m}\ \Sigma\ \text{sign}[\frac{1}{2}-\frac{R_i}{N+1}]$	$\frac{1}{m}\Sigma\ \log\{2[\min(\frac{R_i}{N+1},1-\frac{R_i}{N+1})]\}$
"Ansari-Bradley"	$\frac{3}{m}\ \Sigma\ \min(\frac{R_i}{N+1},1-\frac{R_i}{N+1})\text{sign}(\frac{1}{2}-\frac{R_i}{N+1})$	$3+\frac{12}{m}\Sigma(\frac{R_i}{N+1}-\frac{1}{2})\text{sign}(\frac{1}{2}-\frac{R_i}{N+1})$
"Quartile"	$\frac{3}{m}\left[\sum_{R_i<\frac{1}{4}(N+1)}(\frac{R_i}{N+1})+\sum_{R_i>\frac{3}{4}(N+1)}(\frac{R_i}{N+1})\right]$	$-\frac{1}{m}\sum_{R_i\notin[\frac{1}{4}(N+1),\frac{3}{4}(N+1)]}1$
Exponential	(Not covered by RKHS Theory) (Assumes θ known)	$1+\frac{1}{m}\sum_{i=1}^{m}\log(1-\frac{R_i}{N+1})$

which yields

$$\hat{\theta} = -\frac{1}{m}\sum_{i=1}^{m}\Phi^{-1}\left(\frac{R_i}{N+1}\right)$$

and

$$\hat{\psi} = \frac{1}{2} - \frac{1}{2m}\sum_{i=1}^{m}\left[\Phi^{-1}\left(\frac{R_i}{N+1}\right)\right]^2 .$$

(Logistic) Again, since the tail conditions hold, we have

$$\langle f_oQ_o, f_oQ_o\rangle = \int_0^1 (2u-1)^2 du = \frac{1}{3} .$$

By symmetry of f_o, $\langle f_oQ_o, Q_o(f_oQ_o)\rangle = 0$.

By Eubank (1979),

$$\langle Q_o(f_oQ_o), Q_o(f_oQ_o)\rangle = \int_0^1 \left[1-(2u-1)\log\frac{u}{1-u}\right]^2 du = \frac{3+\pi^2}{9} .$$

Then,

$$\Sigma^{-1} = \begin{bmatrix} 3 & 0 \\ 0 & \frac{9}{3+\pi^2} \end{bmatrix} .$$

For $\underline{g}$, we have

$$\langle f_oQ_o, \tilde{D}(u)-u\rangle = -\int_0^1 (2u-1)\,d[\tilde{D}(u)-u] = 1 - \frac{2}{m}\sum_{i=1}^{m}\left(\frac{R_i}{N+1}\right),$$

and

$$\langle Q_o(f_oQ_o), \tilde{D}(u)-u\rangle = \int_0^1 \left[1-\left(\log\frac{u}{1-u}\right)(2u-1)\right]d[\tilde{D}(u)-u]$$

$$= -\frac{1}{m}\sum_{i=1}^{m} Q_o\left(\frac{R_i}{N+1}\right)J_o\left(\frac{R_i}{N+1}\right) ,$$

which yields

$$\hat{\theta} = 3 - \frac{6}{m} \sum_{i=1}^{m} \left(\frac{R_i}{N+1}\right)$$

and

$$\hat{\psi} = -\left(\frac{9}{3+\pi^2}\right) \frac{1}{m} \sum_{i=1}^{m} \log\left[\left(\frac{R_i}{N+1}\right)/\left(1-\frac{R_i}{N+1}\right)\right]\left[2\frac{R_i}{N+1} - 1\right] .$$

(Cauchy) Since the tail conditions hold, we have

$$\langle f_o Q_o, f_o Q_o \rangle = \int_0^1 \sin^2(2\pi u)\,du = \frac{1}{2} .$$

By symmetry of f_o, $\langle f_o Q_o, Q_o(f_o Q_o) \rangle = 0$

and

$$\langle Q_o(f_o Q_o), Q_o(f_o Q_o) \rangle = \int_0^1 [1 + \sin(2\pi u)\tan\pi(u - \tfrac{1}{2})]^2 du = \frac{5}{2} .$$

Then,

$$\Sigma^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{5} \end{bmatrix} .$$

For $\underline{g}$, we have

$$\langle f_o Q_o, \tilde{D}(u) - u \rangle = \frac{1}{m} \sum_{i=1}^{m} \sin\left[2\pi\left(\frac{R_i}{N+1}\right)\right]$$

and

$$\langle Q_o(f_o Q_o), \tilde{D}(u) - u \rangle = 1 + \frac{1}{m} \sum_{i=1}^{m} \sin\left(2\pi\frac{R_i}{N+1}\right)\tan\left[\pi\left(\frac{R_i}{N+1} - \frac{1}{2}\right)\right] ,$$

which yields

$$\hat{\theta} = \frac{2}{m} \sum_{i=1}^{m} \sin 2\pi\left(\frac{R_i}{N+1}\right) ,$$

and

$$\hat{\psi} = \frac{2}{5} - \frac{2}{5m} \sum_{i=1}^{m} \sin[2\pi(\frac{R_i}{N+1})]\tan[\pi(\frac{R_i}{N+1} - \frac{1}{2})] \quad .$$

(Double Exponential)

$$<f_oQ_o, f_oQ_o> = \int_0^{\frac{1}{2}} 1^2 du + \int_{\frac{1}{2}}^{1}(-1)^2 du + \lim_{p\to 0} \frac{1}{p}[f_oQ_o(p)]^2 + \lim_{q\to 1} \frac{1}{1-q}[f_oQ_o(q)]^2$$

$$= 1 + \lim_{p\to 0} p + \lim_{q\to 1} (1-q) = 1 \ .$$

Again, by symmetry

$$<f_oQ_o, Q_o(f_oQ_o)> = 0 + \lim_{p\to 0} \frac{1}{p} p^2 \log^2 p + \lim_{p\to 0} \frac{-1}{1-q}(1-q)^2 \log 2(1-q) = 0.$$

$$<Q_o(f_oQ_o), Q_o(f_oQ_o)> = \int_0^{\frac{1}{2}} [1+Q_o(u)]^2 du + \int_{\frac{1}{2}}^{1} [1-Q_o(u)]^2 du$$

$$+ \lim_{p\to 0} \frac{1}{p}[p\log 2p]^2 + \lim_{q\to 1} \frac{1}{1-q}[(1-q)\log 2(1-q)]^2$$

$$= 2\int_0^{\frac{1}{2}} (1+\log 2u)^2 du - 2\lim_{p\to 0} p - 2\lim_{p\to 0} p = 1 \quad .$$

Then,

$$\Sigma = \Sigma^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

For $\underline{g}$, we have

$$<f_oQ_o, \tilde{D}(u)-u> = \int_0^{\frac{1}{2}} d[\tilde{D}(u)-u] + \int_{\frac{1}{2}}^{1} (-1)d[\tilde{D}(u)-u] + \lim_{p\to 0} p[\tilde{D}(p)-p]$$

$$- \lim_{q\to 1} (1-q)[\tilde{D}(1-q)-(1-q)]$$

$$= \frac{1}{m} \sum_{i=1}^{m} \text{sign} \left[\frac{1}{2} - \frac{R_i}{N+1}\right] \quad .$$

Next,

$$<Q_o(f_oQ_o),\tilde{D}(u)-u> = \int_0^{1/2}(1+\log 2u)\,d[\tilde{D}(u)-u] + \int_{1/2}^{1}[1+\log 2(1-u)]\,d[\tilde{D}(u)-u]$$

$$+ \lim_{p\to 0} p\log(2p)[\tilde{D}(p)-p]$$

$$+ \lim_{q\to 1}\{(1-q)\log[2(1-p)][\tilde{D}(q)-q]\}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\log 2[\min(\frac{R_i}{N+1}, 1-\frac{R_i}{N+1})]$$

which yields

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\text{sign}\,[\frac{1}{2}-\frac{R_i}{N+1}]$$

and

$$\hat{\psi} = \frac{1}{m}\sum_{i=1}^{m}\log 2[\min(\frac{R_i}{N+1}, 1-\frac{R_i}{N+1})] \quad .$$

("Ansari-Bradley")

$$<f_oQ_o,f_oQ_o> = \int_0^{1/2}(-4u)^2du + \int_{1/2}^{1}[4(1-u)]^2du + \lim_{p\to 0}\frac{1}{p}[2p^2]^2$$

$$+ \lim_{q\to 1}\frac{1}{1-q}[2(1-q)^2]^2$$

$$= \frac{2}{3}+\frac{2}{3} = \frac{4}{3} \quad .$$

Again, by symmetry of f_o,

$$<Q_o(f_oQ_o),f_oQ_o> = 0 + \lim_{p\to 0}\frac{1}{p}(2p^2)^2(1-\frac{1}{2}p)$$

$$+ \lim_{q\to 1}\frac{1}{1-q}[2(1-q)^2]^2[\frac{1}{2}(\frac{1}{1-q}) - 1] = 0 \quad .$$

Now, we have

$$\langle Q_o(f_oQ_o),Q_o(f_oQ_o)\rangle = \int_0^{\frac12}(4u-1)^2du + \int_{\frac12}^1 [4(1-u)-1]^2\,du$$

$$+ \lim_{p\to 0}\frac{1}{p}[(1-\frac{1}{2p})2p^2]^2 + \lim_{q\to 1}\frac{1}{1-q}$$

$$\{[-1+\frac{1}{2}\,\frac{1}{1-q}]2(1-q)^2\}^2$$

$$= \frac{1}{3} + 0 + 0 = \frac{1}{3} \quad .$$

Then,

$$\Sigma^{-1} = \begin{bmatrix} \frac{4}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}^{-1} \quad .$$

For $\underline{g}$, we have

$$\langle f_oQ_o,\tilde{D}(u)-u\rangle = \int_0^{\frac12} 4u\ d[\tilde{D}(u)-u] + \int_{\frac12}^1 [-4(1-u)]d[\tilde{D}(u)-u]$$

$$+ \lim_{p\to 0}\frac{1}{p}[(1-\frac{1}{2p})2p^2][\tilde{D}(p)-p]$$

$$+ \lim_{q\to 1}\frac{1}{1-q}[(\frac{1}{2}\,\frac{1}{1-q} - 1)2(1-q)^2][\tilde{D}(q)-q]$$

$$= \frac{4}{m}\sum_{i=1}^{m} \operatorname{sign}(\frac{1}{2} - \frac{R_i}{N+1})\ \min\ (\frac{R_i}{N+1},\ 1 - \frac{R_i}{N+1}) \quad .$$

Next,

$$<Q_o(f_oQ_o), \tilde{D}(u)-u> = \int_0^1 Q_o(u)J_o(u)du - \int_0^1 Q_o(u)J_o(u)d\,\tilde{D}(u)$$

$$+ \lim_{p\to 0} \frac{1}{p}(1 - \frac{1}{2p})2p^2[\tilde{D}(p)-p]$$

$$+ \lim_{q\to 1} \frac{1}{1-q}(\frac{1}{2(1-q)} -1)2(1-q)^2[\tilde{D}(q)-q]$$

$$= 1 + \frac{4}{m} \sum_{i=1}^{m} \text{sign}(\frac{1}{2} - \frac{R_i}{N+1})(\frac{R_i}{N+1} - \frac{1}{2}) \quad ,$$

which yields

$$\hat{\theta} = \frac{3}{m} \sum_{i=1}^{m} \min\left(\frac{R_i}{N+1}, 1 - \frac{R_i}{N+1}\right) \text{ sign } \left(\frac{1}{2} - \frac{R_i}{N+1}\right)$$

and

$$\hat{\psi} = 3 + \frac{12}{m} \sum_{i=1}^{m} \left(\frac{R_i}{N+1} - \frac{1}{2}\right) \text{ sign } \left(\frac{1}{2} - \frac{R_i}{N+1}\right) \quad .$$

("Quartile")

$$<f_oQ_o, f_oQ_o> = \int_0^1 J_o^{\,2}(u)du + \lim_{p\to 0} \frac{1}{p}(16)^2 \; p^4 + \lim_{q\to 1} \frac{1}{1-q}(16)^2(1-q)^4 = \frac{32}{3}.$$

By symmetry,

$$<f_oQ_o, Q_o(f_oQ_o)> = 0 + \lim_{p\to 0} \frac{1}{p}(16p^2)^2(-\frac{1}{16})\frac{1}{p}$$

$$+ \lim_{q\to 1} \frac{1}{1-q}\,[16(1-q)^2]\,\frac{1}{16}\,\frac{1}{1-q} = 0 \;.$$

Next,

$$<Q_o(f_oQ_o),Q_o(f_oQ_o)> = \int_0^{\frac{1}{4}} [1-(-\frac{1}{16}u^{-1})(-32u)]^2 du$$

$$+ \int_{1/4}^{3/4} \quad [1-(u-\frac{1}{2})0]^2 du$$

$$+ \int_{3/4}^{1} \quad [1-\frac{1}{16}(1-u)^{-1}32(1-u)]^2 du$$

$$+ \lim_{p\to 0} \frac{1}{p}[(-\frac{1}{16})p^{-1}16p^2]^2$$

$$+ \lim_{q\to 1} \frac{1}{1-q} \; [\frac{1}{16}(\frac{1}{1-q})16(1-q)^2]^2$$

$$= \; 1 + 0 + 0 = 1 \quad .$$

Then,

$$\Sigma^{-1} = \begin{bmatrix} \frac{32}{3} & 0 \\ 0 & 1 \end{bmatrix}^{-1} \quad .$$

For $\underline{g}$, we have

$$<f_oQ_o,\tilde{D}(u)-u> = \int_0^1 -J_o(u)d[\tilde{D}(u)-u] + \lim_{p\to 0} \frac{1}{p} \; 16p^2[\tilde{D}(p)-p]$$

$$+ \lim_{q\to 1} \frac{1}{1-q} \; 16(1-q)^2[\tilde{D}(q)-q]$$

$$= \frac{32}{m}\left[\sum_{R_i<\frac{1}{4}(N+1)} (\frac{R_i}{N+1}) + \sum_{R_i>\frac{3}{4}(N+1)} (\frac{R_i}{N+1}-1)\right] \quad .$$

Next,

$$\langle Q_o(f_oQ_o), \tilde{D}(u)-u\rangle = -\int_0^{1/4} d\tilde{D}(u) - \int_{3/4}^{1} d\tilde{D}(u)$$

$$+ \lim_{p\to 0} \frac{1}{p}[\tilde{D}(p)-p](-\frac{1}{16}) \frac{1}{p} 16p^2$$

$$+ \lim_{q\to 1} \frac{1}{1-q}[\tilde{D}(q)-q] \frac{1}{16} \frac{1}{1-q} 16(1-q)^2$$

$$= -\frac{1}{m} \sum_{R_i \notin [\frac{1}{4}(N+1), 3/4(N+1)]} 1 ,$$

which yields

$$\hat{\theta} = \frac{3}{m} \left[\sum_{R_i < \frac{1}{4}(N+1)} (\frac{R_i}{N+1}) + \sum_{R_i > 3/4(N+1)} (\frac{R_i}{N+1} - 1) \right]$$

and

$$\hat{\psi} = -\frac{1}{m} \sum_{R_i \notin [\frac{1}{4}(N+1), 3/4(N+1)]} 1 .$$

(Exponential) Here we use $f_o(x) = e^{-x}$, $x > 0$, and assume $\theta = 0$ since

$$\langle f_oQ_o, f_oQ_o\rangle = \int_0^1 J_o^{\,2}(u)du + \lim_{p\to 0} \frac{1}{p}[f_oQ_o(p)]^2$$

$$+ \lim_{q\to 1} \frac{1}{1-q}[f_oQ_o(q)]^2 = 1 + \infty + 0 = \infty .$$

This means we use the model $D(u)-u = \psi\, Q_o(f_oQ_o)$. We have

$$\langle Q_o(f_oQ_o), Q_o(f_oQ_o)\rangle = \int_0^1 [1-\log(1-u)^{-1}]^2 du + \lim_{p\to 0}\frac{1}{p}[\log(1-p)^{-1}(1-p)]^2$$

$$+ \lim_{q\ 1}\frac{1}{1-q}[\log(1-q)^{-1}(1-q)]^2$$

$$= 1 + 0 + 0 = 1 \quad .$$

Next,

$$\langle Q_o(f_oQ_o), \tilde{D}(u)-u\rangle = \int_0^1 [1-\log(1-u)^{-1}]d\tilde{D}(u) - \int_0^1 [1+\log(1-u)]du$$

$$+ \lim_{p\to 0}\frac{1}{p}(1-p)\log(1-p)^{-1}[\tilde{D}(p)-p]$$

$$+ \lim_{q\to 1}\frac{1}{1-q}(1-q)\log(1-q)^{-1}[\tilde{D}(q)-q]$$

$$= 1 + \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - \frac{R_i}{N+1}\right) \quad .$$

Assuming $\theta = 0$, we have

$$\hat{\psi} = 1 + \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - \frac{R_i}{N+1}\right) \quad .$$

Here, $\sigma_1 = E(X)$ and $\sigma_2 = E(Y)$.

<u>Remark</u>: Since $\psi = \dfrac{\sigma_2-\sigma_1}{\sigma_1}$, we have $\psi + 1 = \dfrac{\sigma_2}{\sigma_1}$. We note $\dfrac{\sigma_2}{\sigma_1}$ is the ratio of scale parameters which is often studied by researchers (for example, the F-test, Siegel-Tukey, Ansari-Bradley, etc. and more recently by Bhattacharyya (1977)). Thus, the estimators and tests given here for ψ also provide results which may be used for

the ratio of scale parameters, $\frac{\sigma_2}{\sigma_1}$.

2.3 Test Statistics and Confidence Intervals

In this section we provide the asymptotic distribution of $\hat{\theta}$, $\hat{\psi}$, and $\hat{D}(u)$ at fixed u given in Theorem 2.2 and 2.3.

2.3.1 Inferences about θ and ψ

Parzen (1980) provides the joint asymptotic distribution of $\hat{\theta}$ and $\hat{\psi}$ in the following result. The proof of this fact is essentially also given in our method of proof in Theorem 2.5.

<u>Theorem 2.4:</u> If f_oQ_o and $Q_o(f_oQ_o)$ are in the RKHS of B(u) with $p = 1 - q = 0$, $F=G$ and $\hat{\theta}$ and $\hat{\psi}$ are as given in *section 2.1, then as* $\lambda_N = \frac{m}{N} \to \lambda_o$ $(0 < \lambda_o < 1)$ and $N \to \infty$, we have

$$\sqrt{N}\begin{pmatrix}\hat{\theta}-\theta\\ \hat{\psi}-\psi\end{pmatrix} \xrightarrow{D} N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \gamma^{-1}\,\Sigma^{-1}\right],$$

where $\gamma = \lambda_o(1-\lambda_o)$ and Σ is as in section 2.1.

<u>Remark</u>: Statisticians often define quantiles with other notations. For example, define $z_{\frac{\alpha}{2}}$ by $Z \sim N(0,1)$ and $P(Z \le z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$. In terms of the quantile function we define $z_{\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$. We also denote Σ^{-1} by $C = (c_{ij})$. Theorem 2.4 gives us the following confidence intervals, regions, and tests of hypotheses.

Note that the confidence regions are proved correct for θ and ψ zero although we may still wish to use them when θ and ψ are moderate.

Corollary 2.1: If f_o is the correct density and D(u) is given by (1.9), then for $\gamma = \lambda_o(1-\lambda_o)$, we have

(1) A $(1 - \alpha)$ 100% asymptotic confidence interval for θ is

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \left(\frac{c_{11}}{N\gamma}\right)^{\frac{1}{2}},$$

(2) A $(1 - \alpha)$ 100% asymptotic confidence interval for ψ is

$$\hat{\psi} \pm z_{\frac{\alpha}{2}} \left(\frac{c_{22}}{N\gamma}\right)^{\frac{1}{2}},$$

(3) A $(1 - \alpha)^2$ 100% joint confidence region for θ and ψ is given for f_o symmetric by

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \left(\frac{c_{11}}{N\gamma}\right)^{\frac{1}{2}}$$

and

$$\hat{\psi} \pm z_{\frac{\alpha}{2}} \left(\frac{c_{22}}{N\gamma}\right)^{\frac{1}{2}}.$$

Corollary 2.2: For D(u) given by (1.9) a test statistic for H_o: $\theta = \psi = 0$ is the quadratic form

$$L = N\gamma \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix}' \Sigma \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix} \xrightarrow{D} \chi^2(2)$$

Remark: $C = \Sigma^{-1} = (c_{ij})$ was calculated for several f_o in the preceding section and c_{11} and c_{22} are given in the following table for convenience.

4. Diagonal Elements of Limiting Covariance Matrix

f_o	c_{11}	c_{22}
Normal	1	1/2
Logistic	3	$9/(3+\pi^2)$
Cauchy	2	2/5
Double Exponential	1	1
"Ansari-Bradley"	3/4	3
"Quartile"	3/32	1
Exponential	(assume known location parameter $\theta = 0$)	1

If one did not trust the D(u) model, then a nonparametric test of H_o: $F = G$ or H_o: $D(u) = u$ may be constructed from the distribution of $\sup \sqrt{N}\,|\tilde{D}(u)-u| \xrightarrow{D} \sup c|B(u)|$. Durbin (1973) gives expressions for the distribution of $\sup|B(u)|$ which suggest test statistics which do not depend on a parametric model for D(u). One may thus use a test based on $\sup\sqrt{N}|\tilde{D}(u)-u|$ as a diagnostic when comparing two samples.

Using $\hat{D}(u)$ and the asymptotic distribution of $\hat{\theta}$ and $\hat{\psi}$ we may find approximate confidence intervals for D(u) when u is fixed and our model for D(u) is correct.

2.3.2 Confidence Intervals for D(u)

From the asymptotic distribution of $\hat{D}(u)$ given in Theorem 2.5 we may obtain confidence intervals for D(u) at specified u. First, we give two useful results.

Lemma 2.3: (Brown (1970), Corollary 3.1) For two square integrable functions $f_1(u)$ and $f_2(u)$ on $0 \le u \le 1$ and in the RKHS of B(u) for $p = 1 - q = 0$,

$$E[\int_0^1 f_1(y)dB(y)\int_0^1 f_2(y)dB(y)] = \int_0^1 [f_1(y) - \int_0^1 f_1(u)du] \cdot [f_2(y) - \int_0^1 f_2(u)du]dy \ .$$

Lemma 2.4: For $f_o\tilde{Q}_o$ and $Q_o(f_oQ_o)$ in the RKHS of B(u) with $p = 1 - q = 0$,

$$W_1(y) = \frac{J_o(y)}{\int_0^1 J_o^2(u)du} \ ,$$

and

$$W_2(y) = \frac{1 - Q_o(y)J_o(y)}{\int_0^1 [1 - Q_o(u)J_o(u)]^2 du} \ ,$$

we have

(1) $\int_0^1 W_1(y)dy = 0$,

(2) $\int_0^1 W_2(y)dy = 0$, and for f_o symmetric also ,

(3) $\int_0^1 W_1(y)W_2(y)dy = 0$.

Proof:

(1) $\int_0^1 W_1(y)dy = [\int_0^1 J_o^2(u)du]^{-1}\int_0^1 J_o(y)dy = 0$,

since $J_o(u) = -J_o(1-u)$.

(2) $$\int_0^1 W_2(y)dy = \{\int_0^1 [1-Q_o(u)J_o(u)]^2 du\}^{-1}\int_0^1 [1-Q_o(y)J_o(y)]dy$$

$$= 0 ,$$

since

$$\int_0^1 [1-Q_o(v)J_o(y)]dy = \int_0^1 [Q_o(u)f_oQ_o(u)]' du$$

$$= \{Q_o(u)f_o[Q_o(u)]\}\Big|_0^1$$

$$= 0 ,$$

since $Q_o(f_oQ_o)$ is in the RKHS of B(u) with $p = 1 - q = 0$.

(3) $$\int_0^1 W_1(y)W_2(y)dy = \frac{\int_0^1 [-J_o(u)][1-Q_o(u)J_o(u)]du}{(\int_0^1 J_o^2(u)du)(\int_0^1 [1-Q_o(u)J_o(u)]^2 du)}$$

$$= 0$$

since $J_o(u) = -J_o(1-u)$ and $Q_o(u) = -Q_o(1-u)$.

In the following theorem we give the asymptotic distribution of $\dot{D}(u)$ under the null hypothesis, H_o: F = G.

Theorem 2.5: If f_o is symmetric, f_oQ_o and $Q_o(f_oQ_o)$ are in the RKHS of B(u) with $p = 1 - q = 0$, the conditions of Theorem 2.1 hold, and F = G, then as $N \to \infty$ such that $\lambda_N = \frac{m}{N} \to \lambda_o (0 < \lambda_o < 1)$, we have

$$\sqrt{N}\,[D(u) - u] \xrightarrow{L} \left(\frac{1-\lambda_o}{\lambda_o}\right)^{1/2} \{f_oQ_o(u)\int_0^1 W_1(y)\,dB(y)$$

$$+ Q_o(u) f_oQ_o(u) \int_0^1 W_2(y)\,dB(y)\}$$

$$= Z_1(u)$$

which we call "the Brownian bridge representation of $\hat{D}(u)$", where $W_1(y)$ and $W_2(y)$ are given in Lemma 2.4.

<u>Proof:</u> By definition of $D(u)$ we have

$$\sqrt{N}[D(u)-u] = \sqrt{N}(1-\lambda)[\hat{\theta}\, f_1(u) + \hat{\psi}\, f_2(u)],$$

where $f_1(u) = f_oQ_o(u)$ and $f_2(u) = Q_o(u)f_oQ_o(u)$. Also (by definition of $\hat{\theta}$ and $\hat{\psi}$)

$$(1-\lambda)\,\hat{\theta} = \int_0^1 W_1(y)\,d[D(y)-y]$$

and

$$(1-\lambda)\,\hat{\psi} = \int_0^1 W_2(y)\,d[D(y)-y],$$

where W_1 and W_2 are defined in Lemma 2.4, and $\hat{\theta}$ and $\hat{\psi}$ are given in Theorem 2.2. This gives

$$\sqrt{N}[D(u)-u] = f_1(u)\int_0^1 W_1(v)\,d[\sqrt{N}(D(v)-v)] + f_2(u)\int_0^1 W_2(y)\,d[\sqrt{N}(D(v)-y)]$$

$$= \int_0^1 [f_1(u)W_1(v) + f_2(u)W_2(v)]\,d[\sqrt{N}(D(v)-v)] \;.$$

Since $f_1(u)W_1(y)+f_2(u)W_2(y)$ is L_2 and $\sqrt{N}(\tilde{D}(y)-y)\xrightarrow{L} c\,B(y)$,

where $c = (\frac{1-\lambda_o}{\lambda_o})^{\frac{1}{2}}$, we have

$$\sqrt{N}[\hat{D}(u)-u] \xrightarrow{L} \int_0^1 [f_1(u)W_1(y)+f_2(u)W_2(y)]\ dcB(y)$$

$$= c\{f_1(u)\int_0^1 W_1(y)dB(y)+f_2(u)\int_0^1 W_2(y)dB(y)\} \ .$$

This gives,

$$\sqrt{N}[\hat{D}(u)-\ u] \xrightarrow{L} (\frac{1-\lambda_o}{\lambda_o})^{\frac{1}{2}}\{f_oQ_o(u)\int_0^1 W_1(y)dB(y)+Q_o(u)f_oQ_o(u)\int_0^1 W_2(y)dB(y).$$

From this representation we are given the asymptotic distribution of $\hat{D}(u)$.

Remark: Although we state the theorem under H_o: F = G, we hope that for $\theta_N = \theta/\sqrt{N}$ and $\psi_N = \psi/\sqrt{N}$, i.e., local alternatives, we may expect a similar result. The argument needed is exemplified in Lepage (1975), Hajék and Šidák (1967), Chapter VI, and in many of their references but complicated here by the error term.

Corollary 2.3: For the assumptions of Theorem 2.5, a $(1 - \alpha)$ 100% asymptotic confidence interval for D(u) is

$$\hat{D}(u) \pm z_{\frac{\alpha}{2}} \{\frac{c^2}{N} f_o{}^2[Q_o(u)][c_{11}+c_{22}Q_o{}^2(u)]\}^{\frac{1}{2}} ,$$

where $c_{11} = \int_0^1 W_1^2(y)dy$ and $c_{22} = \int_0^1 W_2^2(y)dy$.

Proof: We assume the result of Theorem 2.5 to obtain,

$$V[Z_1(u)] = c^2 V\{f_1(u)\int_0^1 W_1(y)dB(y) + f_2(u)\int_0^1 W_2(y)dB(y)\}$$

$$= c^2 f_o^2 [Q_o(u)]\{V[\int_0^1 W_1(y)dB(y)] + Q_o^2(u)$$

$$\cdot V[\int_0^1 W_2(y)dB(y)]\}$$

since $\int_0^1 W_1(y)W_2(y)dy = 0$. Further,

$$V[\int_0^1 W_1(y)dB(y)] = E\{[\int_0^1 W_1(y)dB(y)]^2\}$$

$$= \int_0^1 [W_1(y) - \int_0^1 W_1(u)du]^2 dy$$

$$= \int_0^1 W_1^2(y)dy \ ,$$

since $\int_0^1 W_1(u)du = 0$ and

$$V[\int_0^1 W_2(y)dB(y)] = E[\int_0^1 W_2(y)dB(y)]^2$$

$$= \int_0^1 [W_2(y) - \int_0^1 W_2(u)du]^2 dy$$

$$= \int_0^1 W_2^2(y)dy$$

Note: The values of c_{11} and c_{22} for seven densities are given in Table 4 (p. 41). The same densities have f_oQ_o and Q_o given in Table 3 (p. 29).

The distribution for $\hat{D}(u)$ will be used for evaluating the model of $D(u)$ and for selecting f_o in section 3.

However, we first illustrate estimating θ and ψ with $0 < p \leq u \leq q < 1$ and provide these trimmed estimators for the exponential distribution.

2.4 Truncated Estimation for the Exponential Distribution

In this section we modify the estimation from $0 \leq u \leq 1$ to $0 < p \leq u \leq 1 - p < 1$ for the two parameter exponential distribution.

Since $f_oQ_o(u) = 1 - u$ does not satisfy the left tail condition that $\lim_{p \to 0} \frac{1}{p} (f_oQ_o)^2(p)$ exists, we can not estimate θ using all $0 \leq u \leq 1$. The following theorem implements simultaneous estimation of θ and ψ based on calculating truncated inner products as defined in formula (1.13) with u truncated to the interval $0 < p \leq u \leq 1 - p < 1$.

<u>Theorem 2.6</u>: For the location and scale exponential density and $0 < p \leq u \leq 1 - p < 1$, if we use Parzen's $\tilde{D}(u)-u$ representation, we have

$$(1-\lambda)\begin{pmatrix}\hat{\theta} \\ \hat{\psi}\end{pmatrix} = \Sigma^{-1} \underline{g} \ ,$$

where

$$\Sigma = (\sigma_{ij}) \text{ and } \underline{g} = (g_i) \ ,$$

and

$$\sigma_{11} = \frac{1-p}{p} \ ,$$

$$\sigma_{12} = \sigma_{21} = \frac{p-1}{p} \log (1 - p) ,$$

$$\sigma_{22} = \frac{1-p}{p} \log^2(1-p) + (1-p) \log (1-p) - p \log p \quad ,$$

$$g_1 = \frac{1-p}{p} \tilde{D}(p) + \tilde{D}(1-p) - \frac{1}{m} (\frac{\#R_i}{N+1} \varepsilon[p, 1-p]) - 1 ,$$

and

$$g_2 = \frac{1}{m} (\frac{\#R_i}{N+1} \varepsilon[p, 1-p]) + \frac{1}{m} \sum_{\frac{R_i}{N+1}\varepsilon[p,1-p]} \log(1- \frac{R_i}{N+1}) - \tilde{D}(p)\frac{1-p}{p} \log(1-p)$$

$$-\tilde{D}(1 - p) \log p + \log p ,$$

which are all calculable from the data and p.

Proof: Parzen (1979, 1980) gives reasons for using $0 < p \leq u \leq 1 - p < 1$. Thus, we have

$$\sigma_{11} = \langle f_oQ_o, f_oQ_o \rangle_{p,1-p} = \int_p^{1-p} J_o^{\,2}(u)du + \frac{1}{p}(f_oQ_o)^2(p)$$

$$+ \frac{1}{1-q} (f_oQ_o)^2(q)$$

$$= u\Big|_p^{1-p} + \frac{(1-p)^2}{p}$$

$$= 1 - p + \frac{(1-p)^2}{p}$$

$$= \frac{1-p}{p} \quad .$$

Also,

$$\sigma_{21} = \sigma_{12} = \langle f_o Q_o, Q_o(f_o Q_o)\rangle_{p,1-p} = \int_p^{1-p}(-1)[1-\log(1-u)^{-1}]du$$

$$+ \frac{1}{p}(1-p)^2\log(1-p)^{-1} + \frac{1}{p}p^2\log(p^{-1})$$

$$= \int_{1-p}^{p}\log y\, dy - 1 + 2p$$

$$-\frac{(1-p)^2}{p}\log(1-p) - p\log p$$

$$= -(1+\frac{1-p}{p})(1-p)\log(1-p)$$

$$= -\frac{1-p}{p}\log(1-p) \quad .$$

Next,

$$\sigma_{22} = \langle Q_o(u) f_o Q_o(u),\ Q_o(u) f_o Q_o(u)\rangle$$

$$= \int_p^{1-p}[1-\log(1-u)^{-1}]^2 du + \frac{1}{p}\log^2(1-p)^{-1}(1-p)^2 + \frac{1}{p}(\log^2 p^{-1})p^2$$

$$= \int_p^{1-p}[1+\log(1-u)]^2 du + \frac{(1-p)^2}{p}\log^2(1-p) + p\log^2 p\ ,$$

where

$$\int_p^{1-p}[1+\log(1-u)]^2 du = \int_p^{1-p}(1+\log y)^2 dy = \int_p^{1-p}1du + 2\int_p^{1-p}\log u\, du$$

$$+ \int_p^{1-p}\log^2 u\, du$$

$$= 1-2p + 2[(1-p)\log(1-p)-1+p-p\log p+p]$$

$$+[(1-p)\log^2(1-p)-(1-p)\log(1-p)$$

$$+1-p-p\log^2 p+p\log p-p]$$

$$= (1-p)\log^2(1-p)+(1-p)\log(1-p)-p\log^2 p-p\log p.$$

Combining the above expressions yields

$$\sigma_{22} = \frac{1-p}{p}\log^2(1-p)+(1-p)\log(1-p)-p\log p\ .$$

For $\underline{g}$, we have

$$g_1 = \langle f_o Q_o, \tilde{D}(u)-u\rangle_{p,1-p} = \int_p^{1-p}[-J_o(u)]d[\tilde{D}(u)-u]+\frac{1}{p}f_oQ_o(p)[\tilde{D}(p)-p]$$

$$+\frac{1}{p}f_oQ_o(1-p)[\tilde{D}(1-p)-1+p]$$

$$= \int_p^{1-p}du-\int_p^{1-p}d\tilde{D}(u)+\frac{1-p}{p}[\tilde{D}(p)-p]+\tilde{D}(1-p)-1+p$$

$$= (1-2p)-\frac{1}{m}[\#\frac{R_i}{N+1}\varepsilon[p,1-p]]+\frac{1-p}{p}\tilde{D}(p)-1+p+\tilde{D}(1-p)-1+p$$

$$= \frac{1-p}{p}\tilde{D}(p)+\tilde{D}(1-p)-1-\frac{1}{m}\{\#\frac{R_i}{N+1}\varepsilon[p,1-p]\}\ .$$

Finally,

$$g_2 = \langle Q_o(f_oQ_o), \tilde{D}(u)-u\rangle_{p,1-p}$$

$$= \int_p^{1-p} [1+\log(1-u)]\,d[\tilde{D}(u)-u] + \frac{-1}{p}(1-p)\log(1-p)[\tilde{D}(p)-p]$$

$$+ \frac{-p}{p} \log p\ [\tilde{D}(1-p)-1+p]$$

$$= \frac{1}{m}\ [\#\frac{R_i}{N+1}\varepsilon[p,1-p]] + \frac{1}{m} \sum_{\frac{R_i}{N+1}\ \varepsilon[p,1-p]} \log(1-\frac{R_i}{N+1}) - (1-2p)$$

$$- \int_p^{1-p} \log u\ du - \frac{1-p}{p} \log(1-p)\tilde{D}(p)$$

$$- (\log p)\tilde{D}(1-p) + (1-p)\log(1-p) + (1-p)\log p$$

$$= \frac{1}{m}[\#\frac{R_i}{N+1}\varepsilon[p,1-p]] + \frac{1}{m} \sum_{\frac{R_i}{N+1}\ \varepsilon[p,1-p]} \log(1-\frac{R_i}{N+1}) + \log p - \frac{1-p}{p}\log(1-p)\tilde{D}(p)$$

$$- \log p\ \tilde{D}(1-p)\ .$$

<u>Corollary 2.4</u>: For the assumptions of Theorem 2.6 and with $p < \frac{1}{N}$, the results of Theorem 2.6 for $\underline{g}$ simplify to

$$g_1 = -1 \text{ and } g_2 = 1 + \frac{1}{m} \sum_{\frac{R_i}{N+1}\ \varepsilon[p,1-p]} \log(1 - \frac{R_i}{N+1})\ .$$

<u>Proof:</u> Since $p < \frac{1}{N}$, we have $\frac{1}{N} \leq \frac{R_i}{N+1} \leq \frac{N}{N+1}$ for all i, $\tilde{D}(p)=0$, $\tilde{D}(1-p)=1$, and $(\#\frac{R_i}{N+1}\ \varepsilon[p,1-p]) = m$, thus giving g_1 and g_2 as desired.

One may still use all of the data when using this corollary and its estimates of θ and ψ. The asymptotic approximations are obtained by replacing all $<f_1,f_2>$ with $<f_1,f_2>_{p,q}$ although the distributions may not hold. We also note that the left tail is where the tail condition is not satisfied and that one may desire to use $0 < p < u \leq 1$ as a basis for the estimation with this density. Other than the corollary, we offer no choice for p at this time. A topic of further research is to choose p to minimize a criteria such as variance or mean square error of $\hat{\theta}$ and $\hat{\psi}$.

In section 2.5 we give some remarks on some finite sample size distributions for $\hat{\theta}$ and $\hat{\psi}$ for $0 \leq u \leq 1$.

2.5 Finite Sample Distributions of $\hat{\theta}$ and $\hat{\psi}$

In this section we discuss finite sample size distributions of $\hat{\theta}$ and $\hat{\psi}$. These are obviously needed when n and m are not large. They would also be very useful in seeing how large n and m will need to be in order to use the asymptotic results.

First, under H_o: $F = G$ we know that each possible ordering of the $\{X_i\}$ and $\{Y_i\}$ in the combined sample is equally likely. One may thus enumerate all possible rankings and record $\hat{\theta}$, $\hat{\psi}$, $\hat{D}(u)$, and $\breve{D}(u)$ for each ordering. This yields the complete distributions under H_o.

Under H_a: $\theta \neq 0$ or $\psi \neq 0$ the rankings are not equally likely and the problem is more complex. We must: (a) find it, (b) simulate it, or (c) approximate it. This is a topic for further research.

Another source for the finite sample size distributions of $\hat{\theta}$ and $\hat{\psi}$ arises from the fact that they are often simply linear transformations (which are monotonic) of classical linear rank statistics which often already have finite sample size tables available. For our use we merely take the appropriate linear transformation of the tabled critical values using the following theorem.

Theorem 2.7: If $\hat{\theta}$ (or $\hat{\psi}$) is a linear transformation of $T = \sum_{i=1}^{N} c_i a(R_i)$ and finite sample tables of percentiles for T are available, then these tables easily yield percentile tables for $\hat{\theta}$ (or $\hat{\psi}$).

Proof: Table 1 (p. 9) gives values of a and b for various linear rank test statistics T which have tables available, such that $a\hat{\theta} + b = T$. Clearly,

$$\alpha = P(T \geq t_\alpha) = P(a\hat{\theta} + b \geq t_\alpha)$$

$$= P(\hat{\theta} \geq a^{-1}(t_\alpha - b)) .$$

If $a\hat{\psi} + b = T$, we have

$$\alpha = P(T \geq t_\alpha) = P(\hat{\psi} \geq a^{-1}(t_\alpha - b)) .$$

Theorem 2.8: For f_o symmetric, an approximate α level finite sample size test for the simultaneous H_o: $\theta = \psi = 0$ is given from a size α_1 test of H_o: $\theta = 0$ and size α_2 test of H_o: $\psi = 0$ where $\alpha_1 = \alpha_2 = 1 - \sqrt{1-\alpha}$.

<u>Proof</u>: Under H_o, $P(\text{rej.} H_o: \theta = 0) = \alpha_1 = P(\text{rej.} H_o: \psi = 0) = \alpha_2 = 1 - \sqrt{1 - \alpha}$. Then,

$$P(\text{rej. } \theta = 0 \text{ or rej. } \psi = 0) = 2(1 - \sqrt{1 - \alpha})$$

$$- P(\text{rej. } H_o: \theta = 0 \text{ and rej. } H_o: \psi = 0) .$$

Since f_o is symmetric, $\hat{\theta}$ and $\hat{\psi}$ are asymptotically independent, so

$$P(\text{rej. } H_o: \theta = 0 \text{ and rej. } H_o: \psi = 0) = (1 - \sqrt{1 - \alpha})^2 .$$

Then,

$$P(\text{rej. } H_o: \theta = \psi = 0) = 2 - 2\sqrt{1 - \alpha} - 1 + 2\sqrt{1 - \alpha} - (1 - \alpha)$$

$$= \alpha ,$$

as desired.

These methods for testing H_o: $F=G$ are based upon linear rank statistics, as the test statistics are functions of linear rank statistics with score functions determined by the assumed model f_o. In order to model the data and to obtain more accurate and interpretable tests and estimators, we will develop methods to determine which of several f_o's best fit the data. We begin this development in the next section.

3. MODEL SELECTION

In this section we begin developing criteria to select a model for $D(u)$. For example, in ordinary regression analysis one often makes a choice among models based on R^2 or predictability of the dependent variable or interpretability of the coefficients of the independent variables. Such criteria may be developed for the approach we take in the two sample problem as given below, e.g. Theorem 3.2. In particular, we will develop criteria for determining whether f_o models the data adequately or whether f_o and a location parameter or scale parameter difference adequately models the data, $\tilde{D}(u)$. In this case, the difference between the predicted and the observed values is $\hat{D}(u) - \tilde{D}(u)$, a stochastic process for $u \in (0,1)$. We state the asymptotic distribution theory for $\hat{D}(u) - \tilde{D}(u)$ in section 3.1 and suggest some measures of fit for the various f_o densities in the results of section 3.2. It is the measure of distance between $\hat{D}(u)$ and $\tilde{D}(u)$ that will allow one to choose f_o which best models the data.

3.1 The Asymptotic Distribution of $\hat{D}(u) - \tilde{D}(u)$ under H_o

We develop this distribution as follows:

(1) From Pyke and Shorack (1968) and our Theorem 2.1 we have, under H_o,

$$\sqrt{N}\,[\tilde{D}(u) - u] \xrightarrow{L} \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u).$$

(2) Using section 2.3, Theorem 2.5, we will have, under H_o,

$$\sqrt{N}\ [\hat{D}(u) - u] \xrightarrow{L} Z_1(u),$$

where $Z_1(u)$ is a zero mean normal process

(3) Then, under H_o,

$$\sqrt{N}[\hat{D}(u) - u] - \sqrt{N}\ [\tilde{D}(u) - u] = \sqrt{N}\ [\hat{D}(u) - \tilde{D}(u)],$$

and we will show, under H_o,

$$\sqrt{N}[\hat{D}(u) - \tilde{D}(u)] \xrightarrow{L} Z(u) = Z_1(u) - \left(\frac{1-\lambda_o}{\lambda_o}\right)^{\frac{1}{2}} B(u),$$

where $Z(u)$ is a known 0 mean normal process given f_o.

One way to characterize $Z(u)$ is directly from $\hat{D}(u)$ and $\tilde{D}(u)$. That is, $\hat{D}(u)$ is a functional of $\tilde{D}(u)$ and we know the asymptotic distribution of $\tilde{D}(u)$. Perhaps a more elegant way to study $Z(u)$ is to use the Brownian bridge representation of $\sqrt{N}[\hat{D}(u)-u]$ and $\sqrt{N}[\tilde{D}(u)-u]$ by studying $Z_1(u) - c\ B(u)$. These arguments are illustrated in the following theorem, for F=G and f_o symmetric.

<u>Theorem 3.1</u>: Under the conditions of theorem 2.5, we have

$$\sqrt{N}\ [\hat{D}(u) - \tilde{D}(u)] \xrightarrow{L} Z(u),$$

where $Z(u)$ is a 0 mean normal process with covariance kernel for

$0 < u_1 \leq u_2 < 1$

$$K_Z(u_1,u_2) = (\frac{1-\lambda_o}{\lambda_o})[u_1(1-u_2) - \frac{f_oQ_o(u_1)f_oQ_o(u_2)}{\int_0^1 J_o^2(u)du}$$

$$- \frac{Q_o(u_1)f_oQ_o(u_1)Q_o(u_2)f_oQ_o(u_2)}{\int_0^1[1-Q_o(u)J_o(u)]^2du}] .$$

Proof: Let $c = (\frac{1-\lambda_o}{\lambda_o})^{\frac{1}{2}}$. Then, as in Theorem 2.5,

$$\sqrt{N}[\hat{D}(u)- u] = \int_0^1[f_oQ_o(u)W_1(y)+Q_o(u)f_oQ_o(u)W_2(y)]d[\sqrt{N}(\tilde{D}(y)-y)] \tag{3.1}$$

and,

$$\sqrt{N}[\tilde{D}(u)- u] = \int_0^1 I_u(y \leq u)d[\sqrt{N}(\tilde{D}(y)-y)], \tag{3.2}$$

where

$$I(y \leq u) = 1 , y \leq u ,$$
$$= 0 , y > u .$$

Subtracting (3.2) from (3.1), we have

$$\sqrt{N}[\hat{D}(u)-\tilde{D}(u)]=\int_0^1[f_1(u)W_1(y)+f_2(u)W_2(y)-I_u(y\leq u)]d[\sqrt{N}(\tilde{D}(y)-y)]$$
$$\xrightarrow{L} c\int_0^1[f_1(u)W_1(y)+f_2(u)W_2(y)-I_u(y\leq u)]dB(y)$$
$$=c[f_1(u)\int_0^1 W_1(y)dB(y)+f_2(u)\int_0^1 W_2(y)dB(y)-B(u)] .$$

Thus, the asymptotic mean of $\sqrt{N}[\hat{D}(u) - \tilde{D}(u)]$ is

$$E\{Z(u)\}= cf_1(u)E[\int_0^1 W_1(y)dB(y)]$$

$$+ cf_2(u) E[\int_0^1 W_2(y)dB(y)] - c E[B(u)]$$

$$= 0 ,$$

since $\int_0^1 W_1(v)dv = \int_0^1 W_2(v)dv = 0$. Letting $0 < u_1 \le u_2 < 1$, we have

$$\begin{aligned}
\mathrm{cov}[Z(u_1),Z(u_2)] &= c^2\, E\{[f_1(u_1)\int_0^1 W_1(v)dB(v) \\
&\quad + f_2(u_1)\int_0^1 W_2(v)dB(v)-B(u_1)] \\
&\quad \cdot[f_1(u_2)\int_0^1 W_1(y)dB(y) \\
&\quad + f_2(u_2)\int_0^1 W_2(y)dB(y)-B(u_2)]\} \\
&= c^2\{f_1(u_1)f_1(u_2)E[\int_0^1 W_1(y)dB(y)]^2 \\
&\quad - E[B(u_2)\int_0^1 W_1(y)dB(y)]f_1(u_1) \\
&\quad + f_2(u_1)f_2(u_2)E[\int_0^1 W_2(y)dB(y)]^2 \\
&\quad - E[B(u_2)\int_0^1 W_2(y)dB(y)]f_2(u_1) \\
&\quad - f_1(u_2)E[B(u_1)\int_0^1 W_1(y)dB(y)] \\
&\quad - f_2(u_2)E[B(u_1)\int_0^1 W_2(y)dB(y)]+u_1(1-u_2)\} \\
&= c^2\{\ u_1(1-u_2)+\frac{f_1(u_1)f_1(u_2)}{\int_0^1 J_o^2(u)du}+\frac{f_2(u_1)f_2(u_2)}{\int_0^1[1-Q_o(u)J_o(u)]^2du} \\
&\quad -\frac{f_1(u_1)f_1(u_2)}{\int_0^1 J_o^2(u)du}-\frac{f_2(u_1)f_2(u_2)}{\int_0^1[1-Q_o(u)J_o(u)]^2du} \\
&\quad -\frac{f_1(u_2)f_1(u_1)}{\int_0^1 J_o^2(u)du}-\frac{f_2(u_2)f_2(u_1)}{\int_0^1[1-Q_o(u)J_o(u)]^2du}\ \}
\end{aligned}$$

$$= c^2\{u_1(1-u_2) - \frac{f_1(u_1)f_1(u_2)}{\int_0^1 J_o^2(u)du} - \frac{f_2(u_1)f_2(u_2)}{\int_0^1 [1-Q_o(u)J_o(u)]^2 du}\}$$

$$= K_Z(u_1,u_2),$$

since $f_1(u) = f_oQ_o(u)$ and $f_2(u) = Q_o(u)f_oQ_o(u)$.

Corollary 3.1: For the assumptions of Theorem 3.1, $\underline{u}=(u_1,u_2,\ldots,u_k)'$, $\hat{\underline{D}}(\underline{u})=[\hat{D}(u_1),\ldots,\hat{D}(u_k)]'$, and $\tilde{\underline{D}}(\underline{u})=[\tilde{D}(u_1),\ldots,\tilde{D}(u_k)]'$, we have

$$\sqrt{N}\ [\hat{\underline{D}}(\underline{u}) - \tilde{\underline{D}}(\underline{u})] \xrightarrow{D} N_k\ (\underline{0},\ \Sigma_k)\ ,$$

where $\Sigma_k = (\sigma_{ij})$, $\sigma_{ij} = K_Z(u_i,u_j)$, and N_k denotes the multivariate normal distribution.

Now, for $\underline{u} = (u_1,u_2,\ldots,u_k)'$ we use

$$\sqrt{N}\ [\hat{\underline{D}}(\underline{u}) - \tilde{\underline{D}}(\underline{u})] \xrightarrow{D} N_k(\underline{0},\ \Sigma_k),$$

where $\Sigma_k = (\sigma_{ij})$ as above under the assumptions that f_o is symmetric and θ and ϕ are "small". Note that Σ_k depends on the underlying f_o.

In essence, by studying $\hat{D}(u_i) - \tilde{D}(u_i)$ for $u_i = \frac{i}{N+1}$; $i = 1, \ldots, N$, we study the residuals of a regression model. One may further study the application of classical methods for analysis of residuals in regression analysis to these residuals.

One may then determine the quantiles, u_i, where $\hat{D}(u)$ fits the data well and where it fits the data badly. The main use for the distribution of $\hat{\underline{D}}(\underline{u}) - \tilde{\underline{D}}(\underline{u})$ is as an indicator of how well an underlying f_o density will model the data, $\hat{D}(u)$. We give some methods for determining the fit of $\hat{\underline{D}}(\underline{u})$ to $\tilde{\underline{D}}(\underline{u})$ in the next section.

3.2 Some Measures of Fit

From the distribution theory for $\hat{\underline{D}}(\underline{u}) - \tilde{\underline{D}}(\underline{u})$ we may devise test statistics, whose distributions will provide a test of the location scale model with f_o specified as the underlying family. Computing these values for several hypothesized f_o families will allow us to choose the most appropriate f_o for modelling the data as a local location and scale difference. This also may indicate a location scale model is not viable or that θ and ψ are too large, if we determine $\hat{\underline{D}}(\underline{u})$ does not fit $\tilde{\underline{D}}(\underline{u})$ within chosen limits.

We may accept $\hat{D}(u)$ as an adequate description of the data provided it matches $\tilde{D}(u)$ at $\{u_i,\ i = 1, \ldots, k\}$ where the u_i are fixed at some particular percentiles of interest, in the sense of Eubank (1979). Another set of u_i of interest may be the data points, i.e., $u_i = \frac{i}{N+1}$; $i = 1, \ldots, N$.

Theorem 3.2: If the conditions of Theorem 3.1 hold, then a measure of the fit of $\hat{D}(u)$ to $\tilde{D}(u)$ is

$$\delta_D^2(\underline{u}) = N\gamma[\hat{\underline{D}}(\underline{u})-\tilde{\underline{D}}(\underline{u})]' \textstyle\sum_k^{-1} [\hat{\underline{D}}(\underline{u})-\tilde{\underline{D}}(\underline{u})] \xrightarrow{D} \chi^2(k)$$

where $\chi^2(k)$ denotes the central chi-square distribution with k d.f.

Proof: Assuming Theorem 3.1 results, this is a standard application of the distribution of quadratic forms of normally distributed vectors. The degrees of freedom depends on the number of u_i chosen, i.e. $\{u_i;\ i = 1, \ldots, k\}$.

We may also choose which of the f_o seem to model the data well by calculating Fisher's extension of Mahalanobis' distance as defined in Kshirsagar (1972). That is, we use

$$\delta_{\underline{D}}(\underline{u}) = \{N\gamma[\hat{\underline{D}}(\underline{u})-\tilde{\underline{D}}(\underline{u})]'\textstyle\sum_k^{-1} [\hat{\underline{D}}(\underline{u})-\tilde{\underline{D}}(\underline{u})]\}^{\frac{1}{2}}$$

as a standardized distance measure of $\hat{\underline{D}}(\underline{u})-\tilde{\underline{D}}(\underline{u})$ for each f_o. We choose the f_o which yields the smallest $\delta_{\underline{D}}(\underline{u})$ as that f_o which best models $\tilde{\underline{D}}(\underline{u})$ by $\hat{\underline{D}}(\underline{u})$. Eubank (1979) gives some indication of optimal u_i values to choose for each f_o.

4. DATA ANALYTIC COMPARISONS WITH OTHER APPROACHES

In this section, we analyze three data sets from the literature. The kneecap data in section 4.1 will illustrate what information our methods provide when we fail to reject H_o: $F=G$. The rat data in section 4.2 illustrates a rejection of H_o due mainly to the location difference and provides some interesting comparisons with other methods. The coronary heart disease data sets in section 4.3 also illustrate rejection of H_o, but the fit of the model suggests further analysis. We only analyze the marginal distributions of the two bivariate components of these coronary heart disease data. We would like to thank David Scott for his kindness in sending us a listing of his unpublished coronary heart disease data for analysis and comparison.

4.1 The Kneecap Data in Switzer (1976)

Switzer (1976) analyzes two sample data with his techniques. The data set is given in his Table 1 as right kneecap congruence angles in degrees for 40 male subjects and 40 female subjects. We know the data was supplied by R. G. Miller, but do not know the questions it was gathered to answer. Consequently, any analysis is limited.

Switzer's analysis gives 94.5% confidence bands on $t_o=G^{-1}F$. The figure (Switzer's Figure 1) appears linear, where the bands are not infinite, and suggests a location scale model for the differences between male and female right congruence kneecap angles. Wilk and Gnanadesikan (1968) have named a plot of q versus G^{-1} [F(q)] a Q-Q plot and pointed out that linearity means Y is a location scale transform of X in distribution. Our $\hat{\theta}$ and $\hat{\psi}$ will estimate this relationship. Switzer also gives 94.5% confidence sets for

$$\max_{-10,-5} [t_o(q)-q], \quad \min_{-10,-5} [t_o(q)-q], \text{ and } (-5+10)^{-1}\int_{-10}^{-5} [t_o(u)-u]du$$

as [-5,12], [-7,8], and [-5.9,10.1]. He points out that these were obtained for a limited range and for Smirnov's confidence procedure only. Otherwise finding such sets is much more difficult. Switzer estimates θ in

$$\text{(a) } t_o(\omega) = \omega + \theta \quad \text{(b) } t_o(\omega) = 2\omega + \theta$$

for three different confidence procedures to obtain the results below.

Procedure	Median	Quantile	Smirnov
(a)	-5,7	-5,11	-2,8
(b)	4,16	5,21	12,14

These models, (a) and (b), are special cases of a general location scale difference between X and Y. He regards the short confidence interval for θ in (b), from a Smirnov based procedure, as indicative of 2 being a bad value of the slope to fit these data and suggests fitting general parameters.

Switzer then parametrizes $t_o(\omega)$ as follows:

$$\text{(a) } t_o(\omega) = (1 + \lambda)\ \omega + \theta$$

$$\text{(b) } t_o(\omega) = \omega + \theta/(1 + \lambda\omega),\ \lambda,\theta \geq 0\ .$$

In (a) the treatment effect increases with ω and in (b) the treatment effect decreases as ω increases. Switzer then reports joint confidence intervals for θ and λ in models as given in his Figure 2.

In the Parzen approach we hypothesize a general location scale model and allow $\theta = \frac{\mu_1-\mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$ to be positive or negative obtaining qualitatively similar models to (a) and (b). Since we assume a location and scale model for G and F, we have

$$\begin{aligned}
t_o(\omega) &= G^{-1}F(\omega) \\
&= \mu_2 + \sigma_2\, Q_o[F(\omega)] \\
&= \mu_2 + \sigma_2\, Q_o[F_o(\frac{\omega-\mu_1}{\sigma_1})] \\
&= \mu_2 + \sigma_2\, (\frac{\omega-\mu_1}{\sigma_1}) \\
&= \mu_2 - \frac{\mu_1}{\sigma_1} + \frac{\sigma_2}{\sigma_1}\ \omega \\
&= \mu_2 - \frac{\mu_1}{\sigma_1} + (1 + \psi)\ \omega\ ,
\end{aligned}$$

where

$$\psi = \frac{\sigma_2-\sigma_1}{\sigma_1} \qquad .$$

So, Switzer's θ in model (a) is $\mu_2 - \frac{\mu_1}{\sigma_1}$ and Switzer's λ is our ψ, except that our ψ can be negative also. We add the assumption that the distributions of Y and X are of the same family, F_o.

Our analysis of Switzer's kneecap data yields the following results. A comparison of the quantile functions, $\tilde{Q}_Y$ and $\tilde{Q}_X$, in Figure A suggests $\mu_2 > \mu_1$ and $\sigma_2 < \sigma_1$ or $\theta > 0$, $\psi < 0$. Three choices of f_o yield the following estimates.

f_o	$\hat{\theta}$	$\sqrt{V(\hat{\theta})}$	$\hat{\psi}$	$\sqrt{V(\hat{\psi})}$	p values for H_o: $\theta=\psi=0$
Normal	.206	.233	−.134	.158	.46
Logistic	.400	.387	−.214	.187	.31
Cauchy	.145	.316	−.100	.141	.70

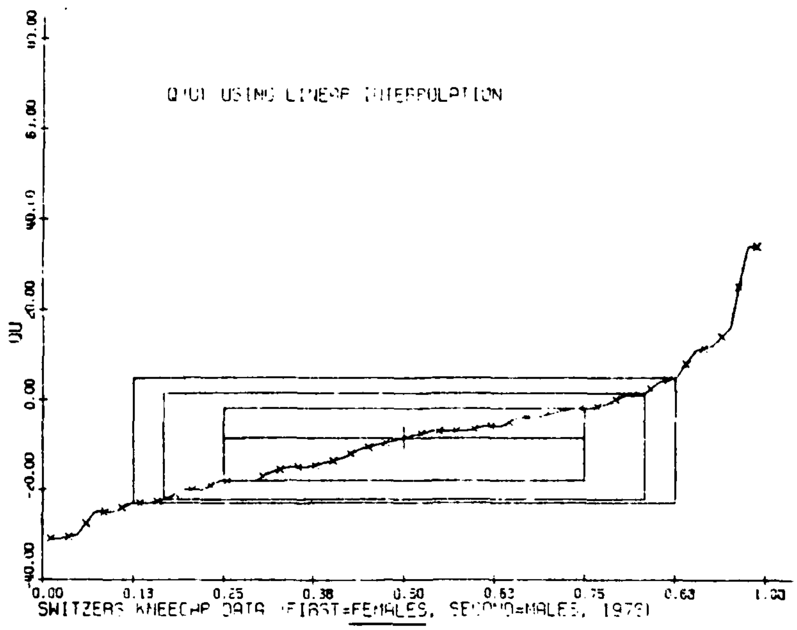
So, regardless of f_o we fail to reject H_o and further remark $\hat{\theta}$ and $\hat{\psi}$ are all within two standard deviations of zero.

A quick comparison of $\hat{D}(u) - \tilde{D}(u)$ graphs in Figure B gives an indication that the logistic density may fit the data best with $\tilde{D}$ rising faster than $\hat{D}$ in all three cases. This indicates $F'/G' = f/g > 1$ at those quantiles. The tests and estimates of section 3 have not been implemented in the computer program yet.

4.2 Doksum and Sievers (1976) Rat Data

These authors have also developed techniques for estimating a general function $t(x)$ where $F(x) = G[\,x + t(x)]$, or $t(x) = G^{-1}F(x) - x$. In fact, Doksum (1974) has developed the asymptotic distribution of $\hat{t}(x) = \tilde{G}^{-1}\tilde{F}(x) - x$. The questions of interest in their paper for the two sample problem are

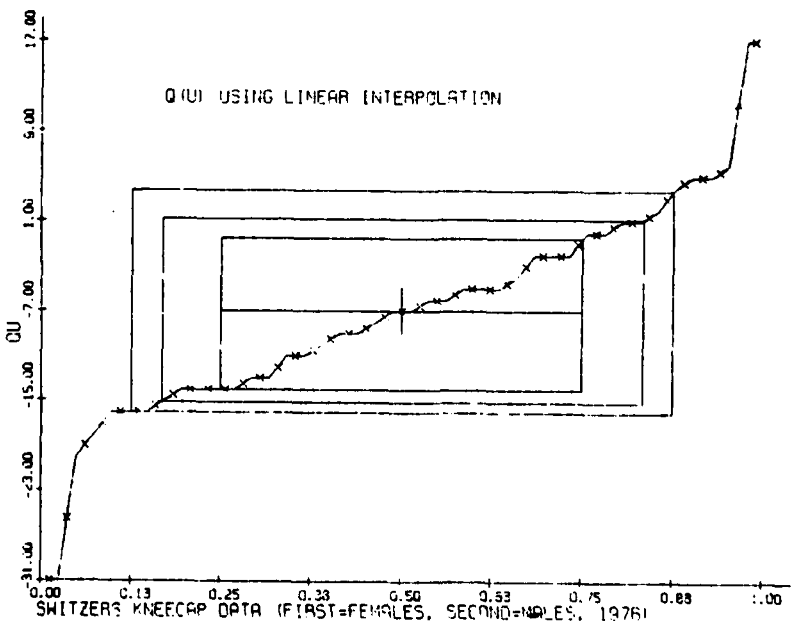
FIGURE A. Quantile Functions for Female and Male Kneecap Data



$\overline{X}_1$ = -8.725

$\tilde{X}_1$ = -8.5

S_1 = 13.2



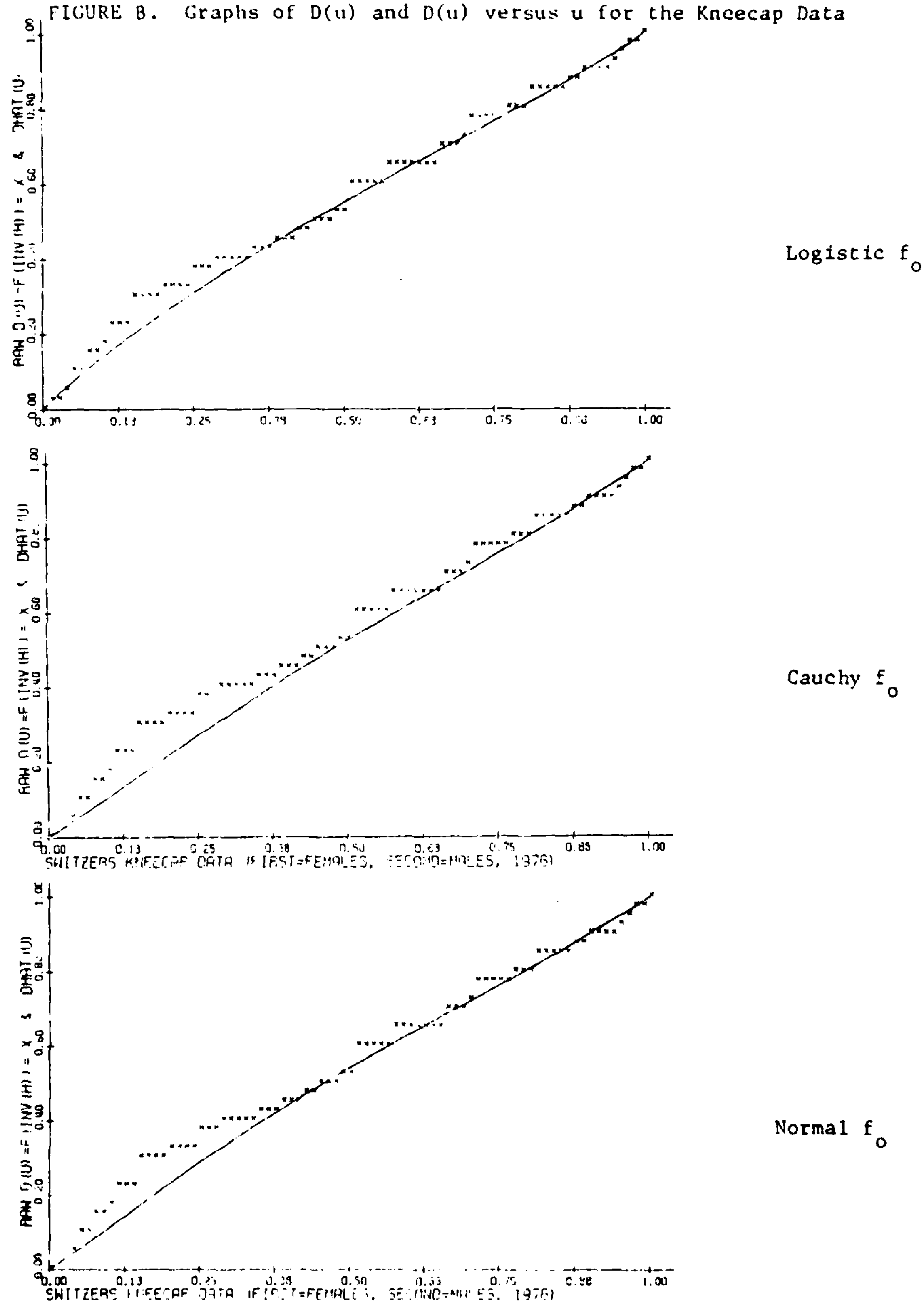
$\overline{X}_2$ = -7.05

$\tilde{X}_2$ = -7

S_2 = 8.843

Note: $(\overline{X}_2-\overline{X}_1)/S_1$ = .13 and $(S_2-S_1)/S_1$ = -.33

FIGURE B. Graphs of D(u) and D̂(u) versus u for the Kneecap Data



(i) Is the treatment beneficial for all the members of the population, i.e., is $t(x) > 0$ for all x?

(ii) If not, for which part of the population is the treatment beneficial, i.e., what is $\{x: t(x) > 0\}$?

(iii) Does a shift model hold, i.e., is $t(x) = \theta$, for some θ and all x?

(iv) If not, does a shift-scale model hold, i.e., is $t(x) = \alpha + \beta x$, for some α and β and for all x ?

All these are answered by giving a confidence band , $[t_*(x), t^*(x)]$, for $t(x)$ simultaneously for all x.

Doksum and Sievers develop "nonparametric" confidence bands by inverting a distribution free Kolmogorov-Smirnov test statistic for H_o: $F=G$. This is their S-band. They give an approximate weighted band (W-band) based on

$$W_N = m^{\frac{1}{2}} \sup \frac{|\tilde{F}(x)-\tilde{G}(x)|}{\psi\{\tilde{H}(x)\}} ,$$

where $\psi(t) = [t(1-t)]^{\frac{1}{2}}$. They remark that this ψ maximizes the minimum power against H_1: $F-G \geq \delta$ for some $\delta > 0$. They give a third nonparametric confidence band (R-band) based on

$$R_N = m^{\frac{1}{2}} \sup \frac{|\tilde{F}(x)-\tilde{G}(x)|}{\tilde{H}(x)} ,$$

the Renyi statistic. The authors present some tables showing why they prefer the W-band to the R-band or S-band except when small quantiles are of interest. Finally, when one is given a location scale model they give a confidence band from order statistics called the O-band. They remark that considerable gain in efficiency is possible with H normal for the O-band over the other bands. In the Parzen approach the Δ_Q model is based on the order statistics, $\tilde{Q}$.

Their example consists of a control group of m = 23 rats and a group of n = 22 rats subjected to one component of California smog, ozone. The weight gain was measured for each rat after seven (7) days in their control or treatment environment. Their Figure 2, a plot of the S-bands, gives six (6) interesting conclusions. They are:

(1) Ozone reduces <u>average</u> weight gain.

(2) Large weight gains are made even larger.

(3) Weight gain is reduced significantly for control weight $x \leq 22.5$.

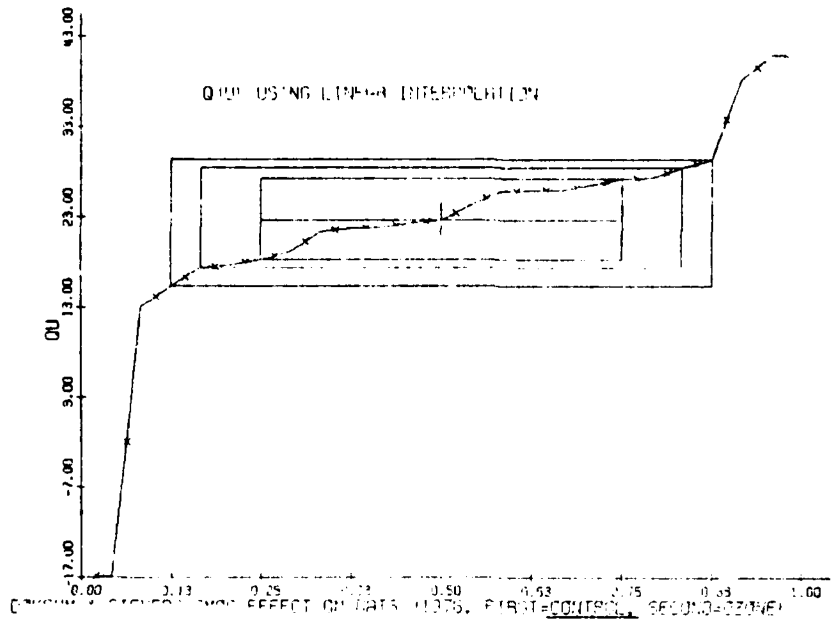
(4) Since a horizontal line fits through the S-bands, we can not reject a shift model.

(5) With a possible outlier left out, $\hat{t}$ appears more linear and thus O-bands could be used which also do not reject a shift model.

(6) They remark that (2) and (3) are strongly suggested and perhaps a larger N would allow the shift model to be rejected.

The Parzen approach suggests differences in scale and location by observation of the two groups quantile box plots (Figure C) when the suspected outlier is included, and a lesser difference in location

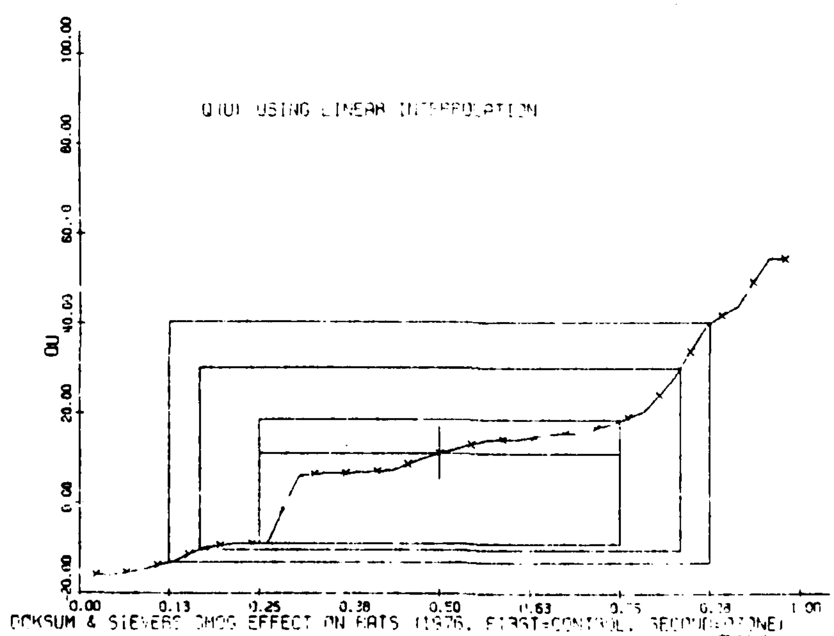
FIGURE C. Quantile Functions for Control and Ozone Rat Data (with outlier)



$\bar{X}_1 = 22.4$

$\tilde{X}_1 = 22.7$

$S_1 = 10.77$



$\bar{X}_2 = 11.01$

$\tilde{X}_2 = 11.1$

$S_2 = 19.02$

Note: $(\bar{X}_2-\bar{X}_1)/S_1 = -1.06$ and $(S_2-S_1)/S_1 = .77$

with no difference in scale when the "outlier" is deleted. Graphical comparisons of the quantile functions for each group suggests a lowering of location from ~22.5 to ~11 with exaggerated loss at lower quantiles and much less exaggerated gain at upper quantiles.

Order Statistics	X_3	X_5	X_8	X_{15}	X_{17}	X_{19}	X_{22}
$\tilde{Q}_Y$	15.4	17.7	21.4	26	26.6	27.4	38.4
$\tilde{Q}_X$	-12.9	-9	6.6	15.5	17.9	28.2	54.6
$\tilde{\Delta}_Q$	28.3	26.7	14.8	10.5	8.7	-.8	-16.2

This leads us to remark that $\hat{t}$ actually levels off at less than 20 at the upper quantiles of x while it goes much below -20 at the lower quantiles of x until the supposed outlier is encountered. With this reinterpretation of their $\hat{t}$ one would agree that the two approaches seem quite consistent, although we do not report a confidence band for Δ_Q in this table, see Theorem 6.2.

Continuing with the Parzen approach, we obtain estimates of $\theta = \frac{\mu_2 - \mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2 - \sigma_1}{\sigma_1}$ for the following f_o families with the "outlier" included:

f_o	$\hat{\theta}$	$\sqrt{V(\hat{\theta})}$	$\hat{\psi}$	$\sqrt{V(\hat{\psi})}$	P Values for H_o: $\theta = \psi = 0$
Normal	- .687	.298	.396	.211	.012
Logistic	-1.531	.517	.503	.249	.002
Cauchy	-1.789	.422	.185	.189	.00008 .

All $\hat{\mu}$ are within $2\hat{\sigma}_{\hat{\mu}}$ of 0 except the logistic which is borderline, and all $\hat{\theta}$ are outside $2\hat{\sigma}_{\hat{\theta}}$ of 0. So, the two samples differ significantly in location and perhaps in scale if we assume the logistic f_o.

Further analysis was done omitting the "outlier" with similar but slightly more revealing results. The quantile functions again suggest more extreme shifts in the lower quantiles.

u	.1	.3	.5	.7	.8	.9
$\tilde{Q}$ Control	15.8	20.5	23.5	26.2	27.4	29.2
$\tilde{Q}$ Treatment	-14.3	0.1	11.1	15.6	19.4	36.8
$\tilde{\Delta}_Q$	-30.1	-20.4	-12.4	-10.6	-8.0	7.6

We also remark that if the -16.9 of the control group was an outlier then perhaps the 54.6 of the ozone group is an "outlier" also. One might conjecture by throwing out more of the tail behavior in these data that the normal D would be the best fitting D(u) model. Examining the $\hat{D}(u)$-$\tilde{D}(u)$ graphs (Figures D/E), seems to indicate the Cauchy f_o does well with the outlier in and the logistic f_o does well with the outlier left out. We would rather accept the extreme behavior of these rats weight gains unless some explanation could be given as to the cause of an error in the measurements resulting in an outlier. We also report the estimates of θ and μ with the "outlier" left out. Then, $(\overline{X}_2-\overline{X}_1)/S_1 = (11.01 - 24.19)/6.68 = -1.97$ and $(S_2-S_1)/S_1 = (19.02 - 6.68)/6.68 = 1.35$

FIGURE D. Graphs of D(u) and D(u) for Rat Data with the Outlier

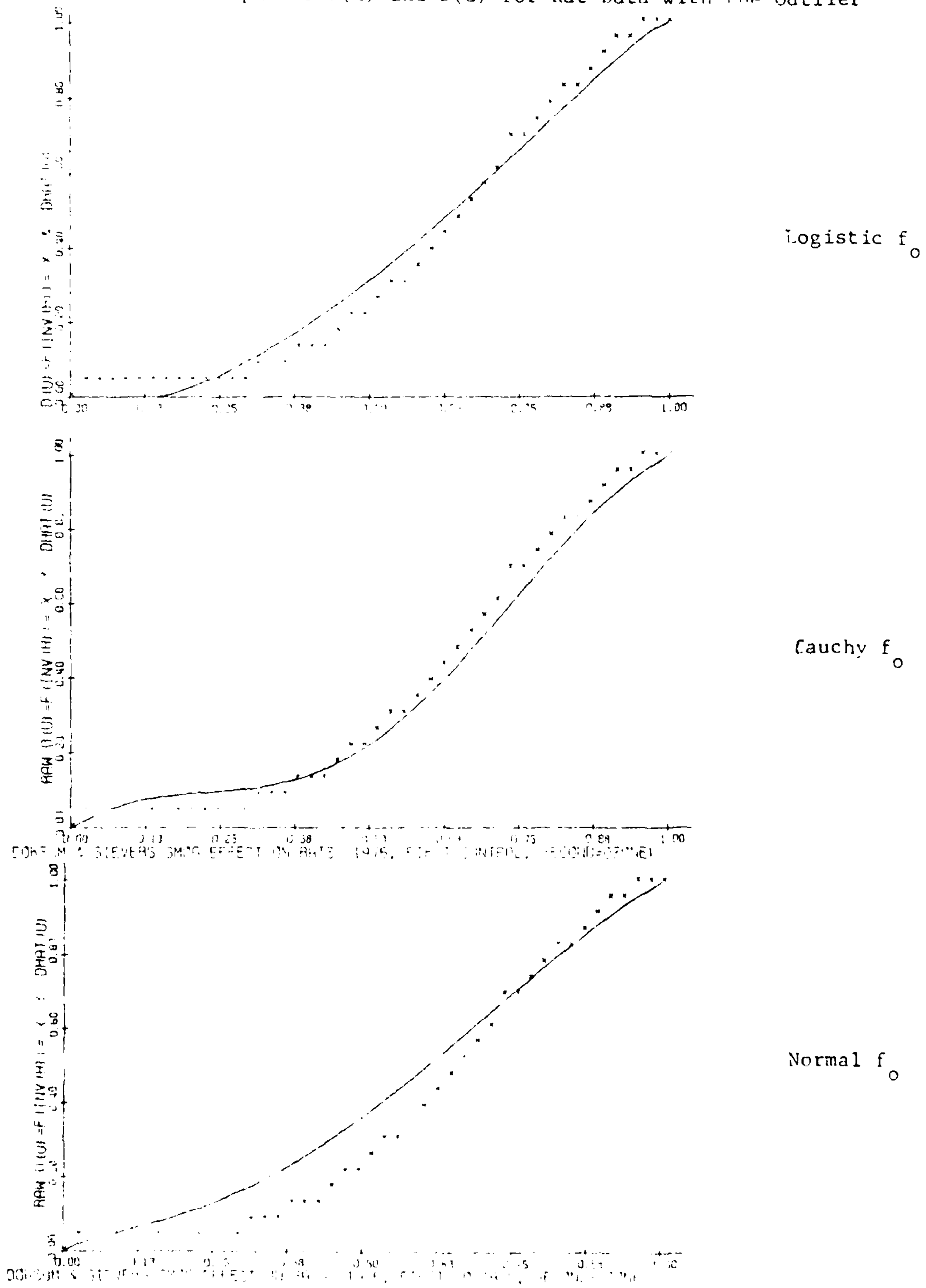
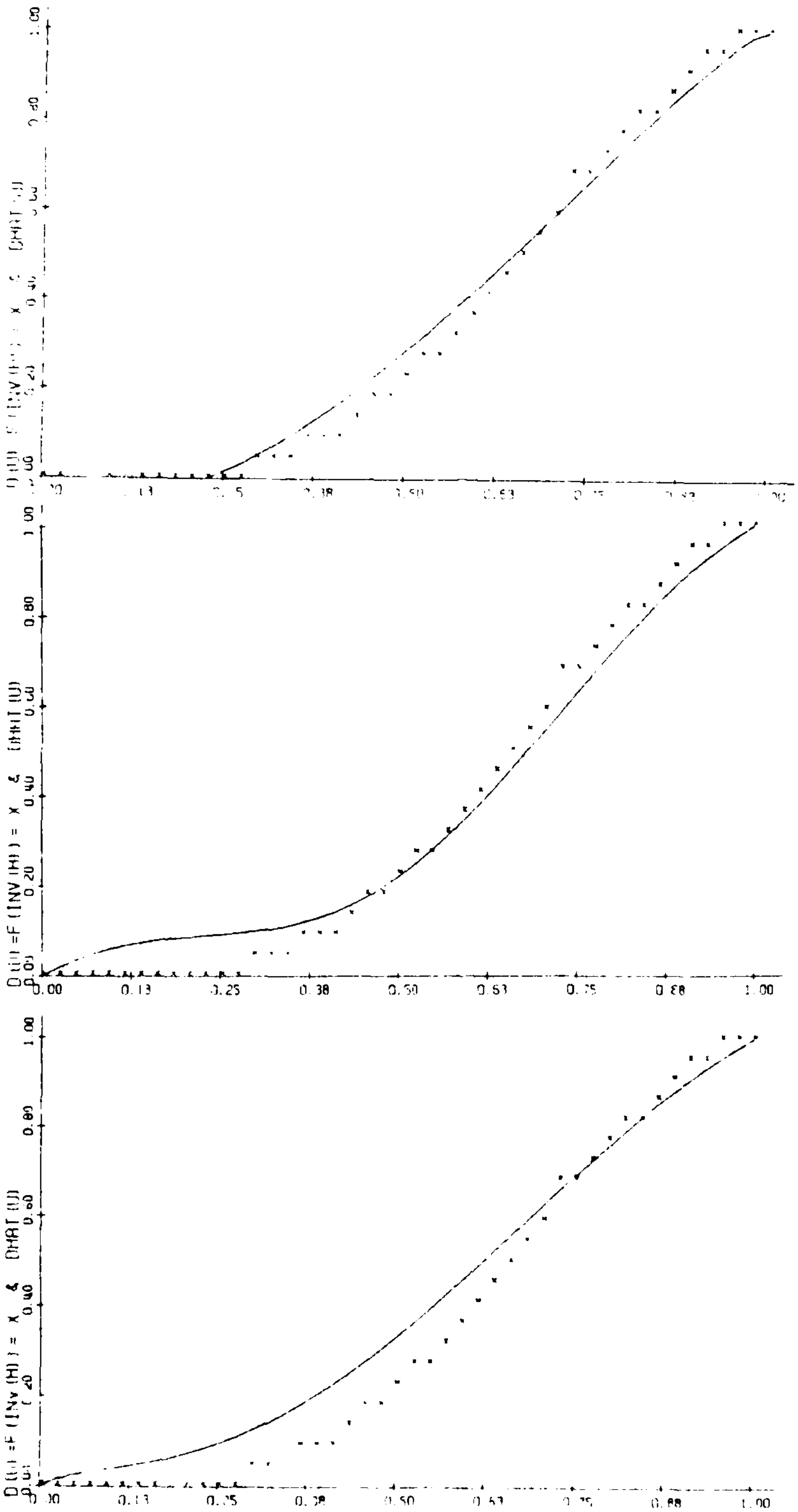


FIGURE E. Graphs of D(u) and $\hat{D}(u)$ for Rat Data Without the Outlier



Logistic f_o

Cauchy f_o

Normal f_o

F_o	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\psi}$	$\sigma_{\hat{\psi}}$	p values for H_o: $\theta=\psi=0$
Normal	-.841	.30	.556	.213	.0007
Logistic	-1.73	.52	.699	.25	.00009
Cauchy	-1.75	.43	.234	.19	.0001

Now, without the one data point all $\hat{\theta}$ are beyond $2\sigma_{\hat{\theta}}$ of 0 and only the Cauchy $\hat{\psi}$ is within $2\sigma_{\hat{\psi}}$. Either the data are distributed Cauchy and the samples differ in location or the data differ in location and scale and are distributed logistic or normal or some unknown other possibility. In either case it appears the location difference is dominant, since $|\hat{\theta}| > |\hat{\psi}|$.

We also note that deleting the one possible outlier did not affect the $\sigma_{\hat{\theta}}$ or $\sigma_{\hat{\psi}}$ very much, but did affect the logistic and normal $\hat{\theta}$ and $\hat{\psi}$. This points out the robustness of the Cauchy model.

4.3 Coronary Heart Disease Data in Scott, et al. (1978)

David Scott of Rice University presented a seminar at Texas A&M where he analyzed, for two groups of patients, measures of plasma tryglycerides and cholesterol. The aim of our analysis is to examine their relation to coronary heart disease. In the control group we have m = 51 patients with no history of coronary heart disease and in the treatment group we have n = 320 patients with a history of coronary heart disease (C.H.D.). The question is "How do the two groups differ in tryglycerides and cholesterol levels?"

Scott's analysis estimated the bivariate density functions of each group and graphically compared them. Although there is a Parzen bivariate quantile approach in the making, we only analyze the marginal quantile functions of the two groups at this time.

First, we examine the tryglycerides. Both groups have similarly shaped quantile functions (Figure F) indicating the distributions may be skewed to the right. The C.H.D. group's tryglycerides tend to be higher but also spread over the non C.H.D. group's tryglycerides for approximately the lower quartile.

On examining $\hat{\theta}$ and $\hat{\psi}$ we see the predominant difference is clearly a shift in location rather than scale.

f_o	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\psi}$	$\sigma_{\hat{\psi}}$	p values for H_o: $\theta=\psi=0$
Normal	.441	.15	.006	.11	.014
Logistic	.797	.26	-.004	.13	.009
Cauchy	.431	.21	-.013	.09	.013

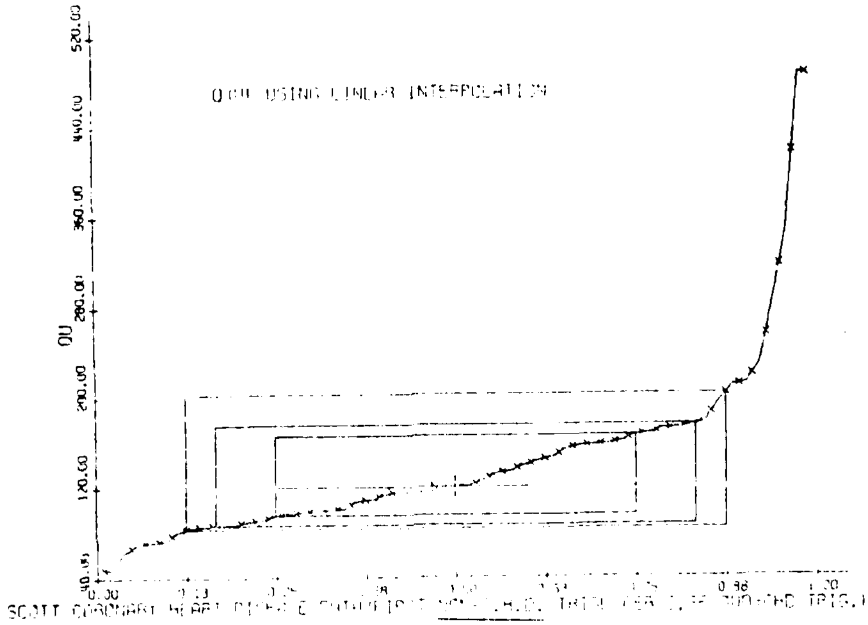
Here $\hat{D}(u)$ for the logistic seems to match $\tilde{D}(u)$ the best (Figure G).

The descriptive $\tilde{\Delta}_Q$ again suggests a skewed distribution for f_o, since $\tilde{\Delta}_Q(u)$ increases with u and $\psi \approx 0$.

u	.125	.25	.5	.75	.875
$\tilde{Q}$ for non CHD	82	91	120	160	195
$\tilde{Q}$ for CHD	91	115	150	218	284
$\tilde{\Delta}_Q$	9	24	30	58	89

Since the quantile functions give an indication of skewness, there should be an f_o which would give more efficient estimates of θ and ψ.

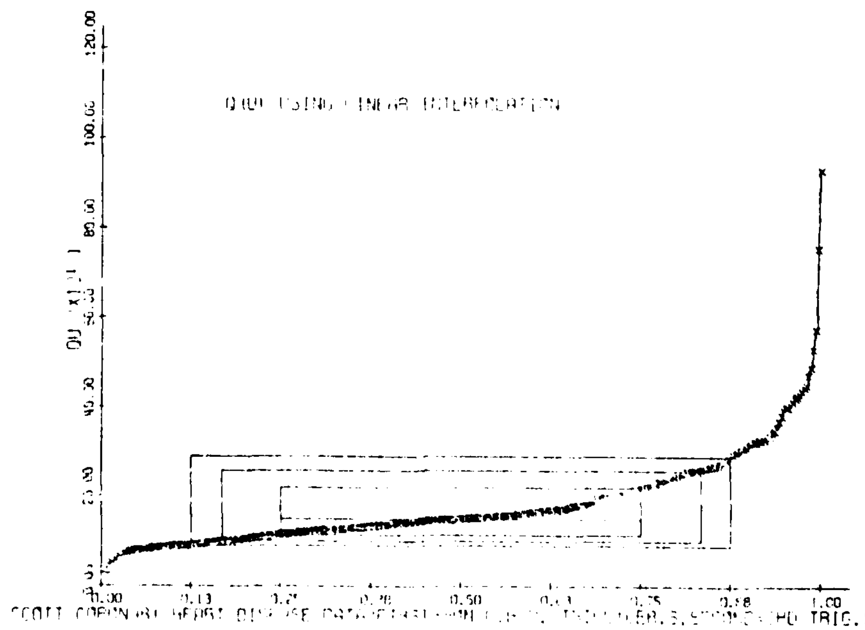
FIGURE F. Quantile Functions for Triglycerides Data



$\bar{X}_1 = 140.35$

$\tilde{X}_1 = 120$

$S_1 = 74.29$



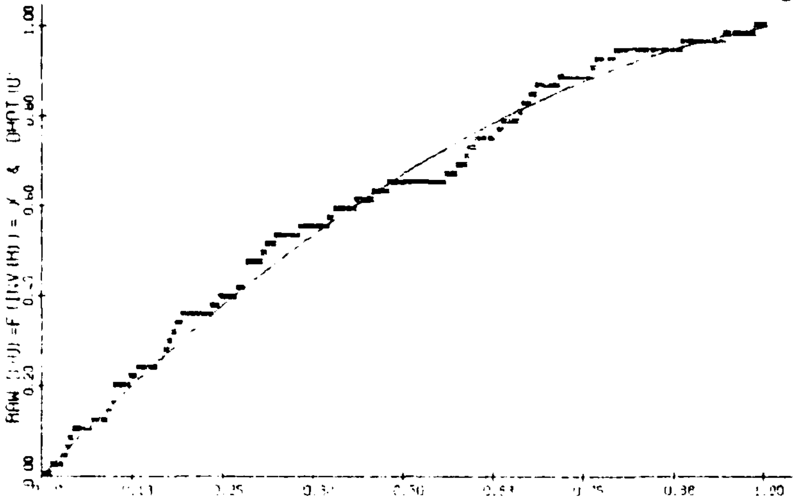
$\bar{X}_2 = 179.04$

$\tilde{X}_2 = 150$

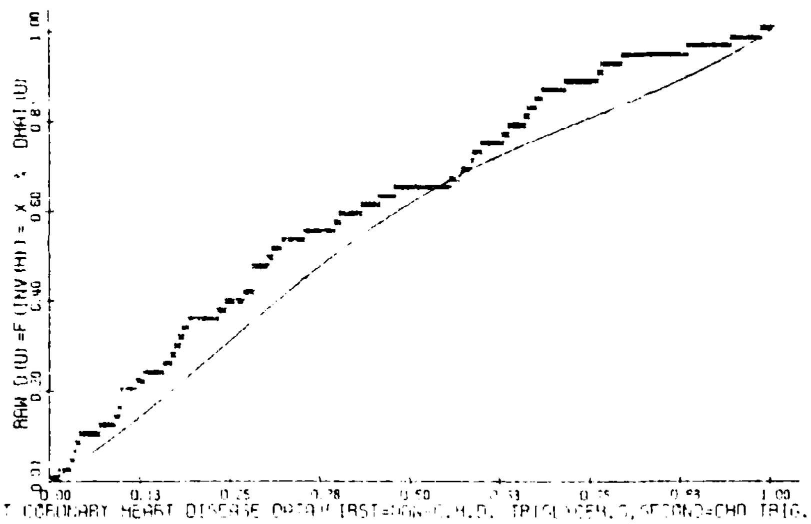
$S_2 = 102.14$

Note: $(\bar{X}_2-\bar{X}_1)/S_1 = .52$ and $(S_2-S_1)/S_1 = .37$

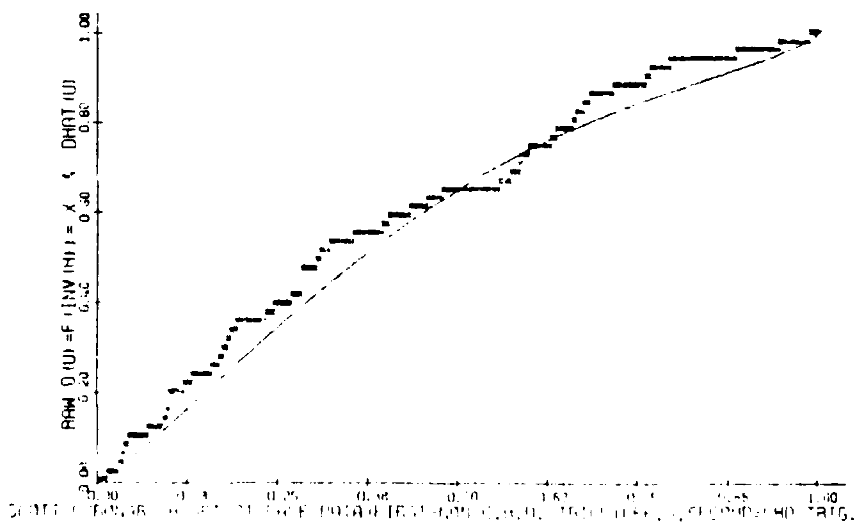
FIGURE C. Graphs of D(u) and $\hat{D}$(u) Versus u for Triglycerides Data



Logistic f_o



Cauchy f_o



Normal f_o

We now examine the marginal distributions of cholesterol levels for each group. The quantile functions (Figure H) are again similar in shape but the C.H.D. group does have a longer tail suggested and is shifted higher suggesting $\theta > 0$.

f_o	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\psi}$	$\sigma_{\hat{\psi}}$	p values for H_o: $\theta=\psi=0$
Normal	.51	.15	-.03	.11	.003
Logistic	.87	.26	-.05	.13	.004
Cauchy	.44	.21	-.007	.10	.117

We conclude cholesterol levels differ in location only, regardless of which of the three f_o are assumed. For this variable $\tilde{\Delta}_Q$ is much more consistent.

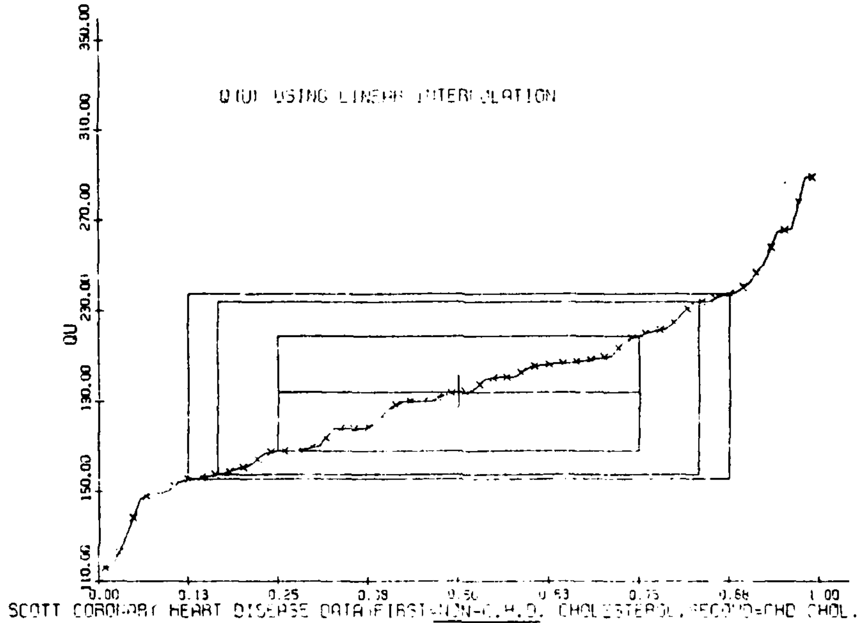
u	.1	.2	.3	.7	.8	.9
$\tilde{Q}$ C.H.D.	168	180	191	236	248	267
non C.H.D.	150.5	161	169.5	208.5	222	239
$\tilde{\Delta}_Q$	17.5	19	21.5	27.5	26	28

We also remark that the graphs of $\hat{D}(u)-\tilde{D}(u)$ (Figure I) suggest the logistic f_o may model these data well.

4.4 Remarks on Examples

From these three example data sets we see that the quantile approach agrees with results of other authors' nonparametric techniques. We agree in accepting H_o as in Switzer's data (section 4.1). We also agree in rejecting H_o as in Doksum and Sievers' data

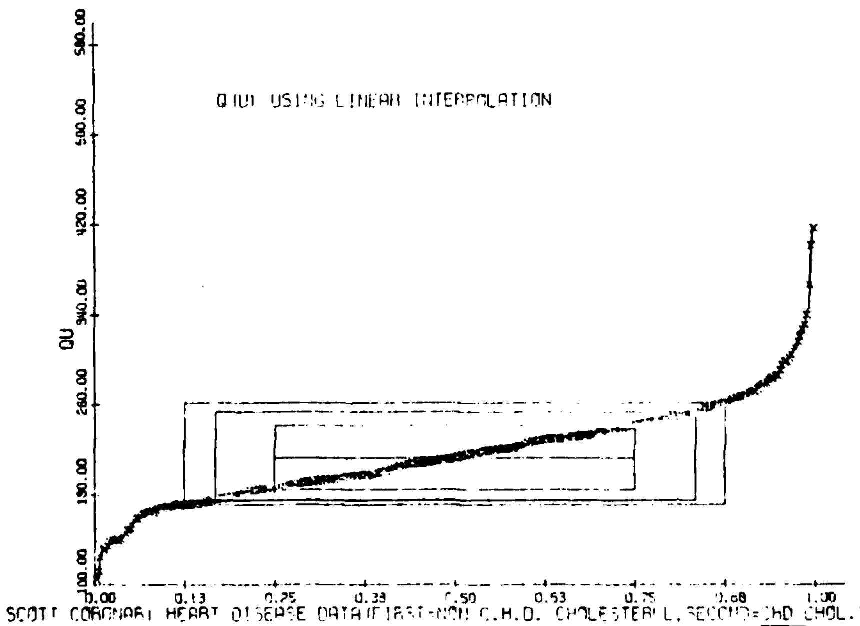
FIGURE H. Quantile Functions for Cholesterol Data



$\overline{X}_1 = 195.16$

$\tilde{X}_1 = 194$

$S_1 = 36.21$



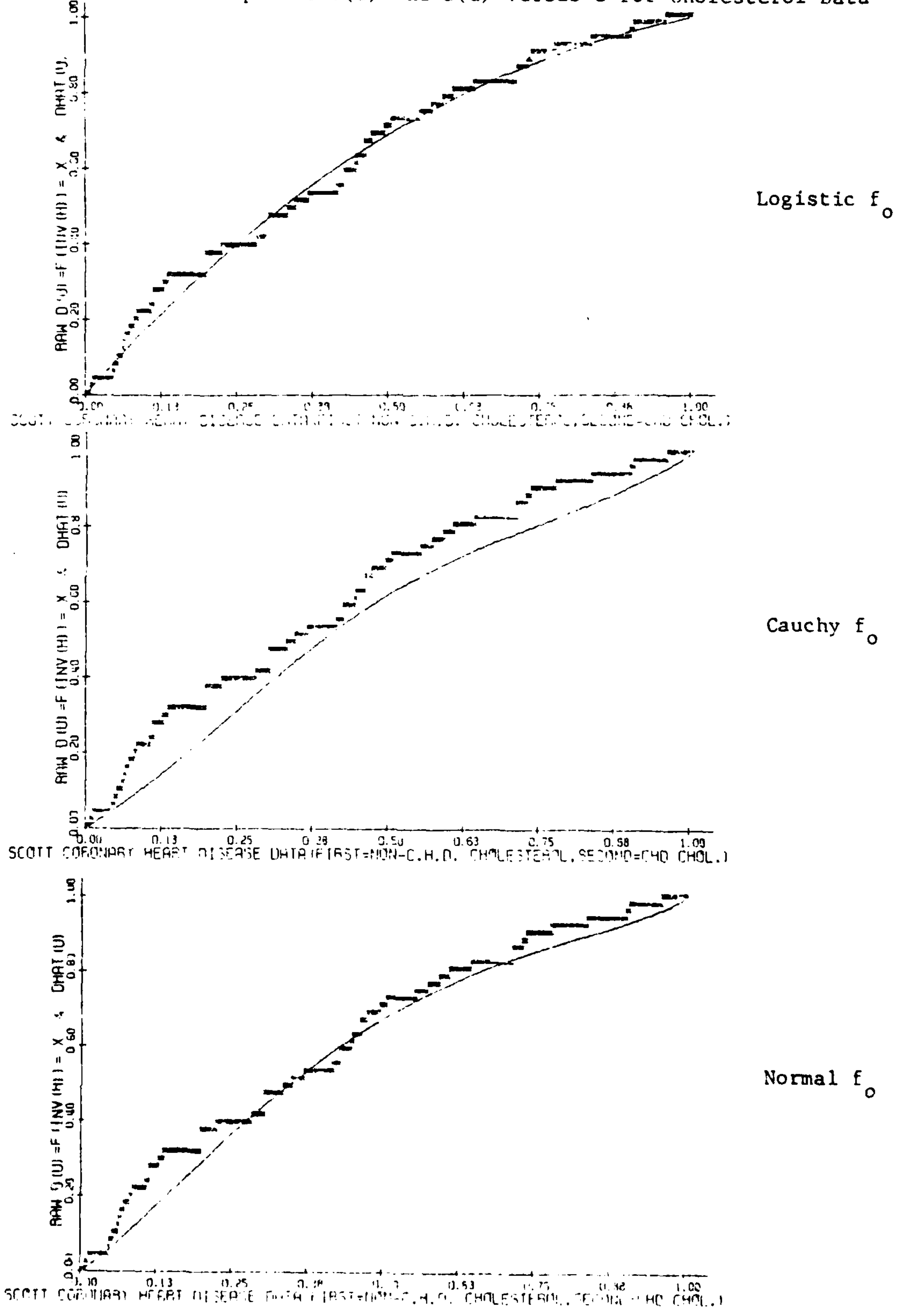
$\overline{X}_2 = 216.2$

$\tilde{X}_2 = 212.5$

$S_2 = 43.01$

Note: $(\overline{X}_2-\overline{X}_1)/S_1 = .58$ and $(S_2-S_1)/S_1 = .19$

FIGURE I. Graphs of D(u) and $\hat{D}(u)$ Versus u for Cholesterol Data



(section 4.2). Also in section 4.2, Parzen's technique seems to indicate the difference of the two groups is best explained by both location and scale shifts rather than just a location shift as it seems Doksum and Sievers may have believed when the supposed outlier is thrown out. When the data point is kept in we best fit the differences with the Cauchy f_o and a location shift similar to Doksum and Sievers results. At any rate the two approaches partially agree and partially disagree. The Parzen approach also models the data and gives some alternate explanation of what may be going on in these two groups of data. This points out how crucial the assumption of that data point being an outlier can be. And finally, in section 4.3 (David Scott's data) we see we can reject H_o with various null families quite consistently, yet be led through the graphical techniques to alternate explanations beyond the analysis performed. That is, we are led to consider some f_o for our $\hat{D}$ estimator which are skewed. Although our approach may now test H_o: $\theta = \psi = 0$ or $F = G$ and estimate θ and ψ for several f_o, some skewed or short tailed densities would also be of interest in modelling some data.

5. OVERVIEW OF THE LITERATURE ON NONPARAMETRIC ESTIMATION AND TESTING OF LOCATION AND SCALE PARAMETERS

The nonparametric estimation of location parameters was started by Hodges and Lehmann (1963). Sen (1966) has extended this technique to scale parameters.

In the decade of the 1970's researchers developed simultaneous estimates of both location and scale parameters. This section reviews the relation of some widely used location and scale tests with estimators in the location and scale model for D(u).

5.1 Location Tests

5.1.1 Linear Rank Tests and $\hat{\theta}$

Linear rank statistics are of the form

$$S = \sum_{i=1}^{N} a(i, R_{Ni}) ,$$

where a is an arbitrary function of i and R_{Ni}, is a relative rank of the X sample. S is a simple linear rank statistic if

$$S = \sum_{i=1}^{N} c_i \, a(R_{Ni}) .$$

Many of the statistics for the two sample problem that have been developed are simple linear rank statistics.

(i) The Van der Waerden test statistic is

$$T_1 = \sum_{i=1}^{m} J_o\left(\frac{R_i}{N+1}\right) = -m\hat{\theta},$$

where $J_o(u) = \Phi^{-1}(u)$ and $\hat{\theta}$ is based on the N(0,1) f_o. This was developed by Van der Waerden (1952) and is asymptotically equivalent to the Fisher-Yates-Terry-Hoeffding normal scores test where $J_o(u_i)$ is replaced by $E[J_o(u_i)]$, Hajek and Šidak (1967).

(ii) The Wilcoxon test statistic [Wilcoxon (1945), Mann and Whitney (1947)] is

$$T_2 = \sum_{i=1}^{m} R_i = 2m - \frac{m(N+1)}{6}\hat{\theta},$$

for $\hat{\theta}$ based on the logistic f_o and Q_o given in Parzen (1979).

(iii) The median test developed by Mood (1950), Westenberg (1948), or Mathisen (1943) is

$$T_3 = \sum_{i=1}^{m} \operatorname{sign}[R_i - \tfrac{1}{2}(N+1)] = m\hat{\theta},$$

for $\hat{\theta}$ based on the double exponential f_o and Q_o.

5.1.2 Exceedance Tests for Location

These tests obtain their name from the fact that they are based on the count of one sample's points which are either above or below the other sample's maximum or minimum value.

They are rather special tests not ordinarily used in a standard analysis. The following are taken from Hájek and Šidák (1967):

(i) The Haga (1959) test with work by Šidák and Vondráček (1957) is based on four quantities: $A = \#$ of $X_i > \max Y_j$, $A' = \#$ of $Y_j > \max X_i$, $B' = \#$ of $X_i < \min Y_j$, and $B = \#$ of $Y_j < \min X_i$ $(i=1, \ldots, m;\ j=1, \ldots, n)$. Then the test statistic

$$T_1 = A + B - A' - B'$$

is optimal under special conditions for the uniform F_o where there is neither an optimal rank test defined or a $\hat{\theta}$ test unless we consider using $\langle f_1, f_2 \rangle_{p,q}$, $0 < p < q < 1$. However, the four quantities A, A', B and B' are related to various comparison functions or $\tilde{D}$ and we mention the exceedance tests to show how they may fit into the general approach taken here. When $\hat{D}(u)$ does not fit $\tilde{D}(u)$ well we may use the Haga test as it is related to various $\tilde{D}(u)$. For $\tilde{D}_1(u) = \tilde{F}\tilde{G}^{-1}(u)$ (proportion of X's $\leq$ Y) we know $A = m$[size of last jump in $\tilde{D}_1(u)$]; $B' = m$[size of first jump in $\tilde{D}_1(u)$]. Similarly, define A' and B for $\tilde{D}_2(u) = \tilde{G}\tilde{F}^{-1}(u)$. Thus, the Haga test is related to first and last jump sizes in the two comparison functions $\tilde{D}_1(u) = \tilde{F}\tilde{G}^{-1}(u)$ and $\tilde{D}_2(u) = \tilde{G}\tilde{F}^{-1}(u)$.

(ii) Rosenbaum's (1954) test is a simpler test designed for the alternative that Y is shifted to the right, $\theta > 0$. In our notation this test statistic is $m[1-\tilde{D}(u \text{ for } \max Y_j)]$ and is more easily adapted to the $\tilde{D}(u)$ used here.

5.1.3 Goodness of Fit Tests for Location

These tests are based on some measure of the distance between $\tilde{F}(x)$ and $\tilde{G}(x)$. We present them here to show their relation to $\tilde{D}(u)$, and thus provide a wider statistical base for the importance of $\tilde{D}(u)$ and $\hat{D}(u)$.

(i) The unnormalized Kolmogorov-Smirnov test statistic is $T = \max_x |\tilde{F}(x) - \tilde{G}(x)|$. Kolmogorov (1933) developed this for a one-sample test and Smirnov (1939) for the two sample test. For $\tilde{D}(u) = \tilde{F}\tilde{H}^{-1}(u)$, we have

$$T = \frac{1}{1-\lambda} \sup |\tilde{D}(u) - u| \quad .$$

Durbin (1973) gives a derivation of the distribution of $\sup|B(u)|$ which may be used for studying the distribution of T. Graphically we plot $\tilde{D}(u) - u$ versus u and see if it significantly exceeds 0 in absolute value which is determined by a given critical value from the distribution of $\sup |B(u)|$.

(ii) The Renyi test is also related to the comparison function and weighted by $\tilde{H}(x)$. It is

$$T_a = \max_A \frac{(n+m)|\tilde{F}(x) - \tilde{G}(x)|}{m\tilde{F}(x) + n\,\tilde{G}(x)} \quad ,$$

where $A = \{x : (n+m)^{-1}[m\tilde{F}(x) + n\,\tilde{G}(x)] = \tilde{H}(x) \geq a\}$. Therefore,

$$T_a = \max_A \left[\frac{|\tilde{D}(u) - u|}{\tilde{H}(x)}\right] \quad .$$

(iii) The Cramér Von Mises test is related as follows

$$T = (n+m)\lambda(1-\lambda)\int_{-\infty}^{\infty}[\tilde{G}(x)-\tilde{F}(x)]^2 d[\frac{m\tilde{F}(x)+n\ \tilde{G}(x)}{m+n}]$$

$$= (n+m)\lambda(1-\lambda)\int_0^1[\tilde{D}(u)-u]^2\ du$$

where $u = \tilde{H}(x)$.

(iv) Finally, we also may remark that Weiss (1976) gives an analogy which shows a two-sample test of H_0: $D(u) - u = 0$ can be developed from any one sample goodness of fit test. Also, Pettitt (1976) gives a two sample Anderson-Darling statistic

$$A^2_{nm} = \frac{nm}{n+m}\int_{-\infty}^{\infty}\frac{(\tilde{F}-\tilde{G})^2}{\tilde{H}(1-\tilde{H})}\ d\tilde{H} = \frac{1}{mn}\sum_{i=1}^{n+m-1}\frac{[M_i(n+m)-n_i]^2}{i(n+m-i)} \quad ,$$

where $M_i = n\tilde{D}\ (\frac{i}{n+m})$. Similar to 2.1.3 (iii) we obtain

$$A^2_{nm} = nm\int_0^1\frac{[\tilde{D}(u)-u]^2}{u(1-u)}\ du \quad ,$$

where $u = \tilde{H}(x)$.

The point we can make with these goodness of fit tests is that they are all functions of $\tilde{D}(u) - u$. They can be computed from the comparison functions and all measure the "size" of $\tilde{D}(u) - u$. Parzen's $\hat{D}(u)$, as well as its extensions, attempts to model $\tilde{D}(u)$ and we will want to minimize the "distance" between $\hat{D}(u)$ and $\tilde{D}(u)$. In other words, we want $\hat{D}(u)$ to converge to the truth so that our estimators, $\hat{\theta}$ and $\hat{\psi}$, are consistent. Other location tests and estimates are given in Table 3 (p. 29).

5.2 Scale Tests

5.2.1 Linear Rank tests and $\hat{\psi}$

These tests are of the same form as in section 5.1, i.e., $S = \sum_{i=1}^{N} a(i, R_{Ni})$ or $\sum_{i=1}^{N} c_i\, a(R_{Ni})$. However, the score functions $a(i,R_{Ni})$ or $a(R_{Ni})$ are different in that they are devised to detect differences in the dispersion or scale parameter of the two distribution functions F and G.

(i) The Klotz (1962) test is

$$T_1 = \sum_{i=1}^{m} [J_o(\frac{R_i}{N+1})]^2 = -2m\,(\hat{\theta} - \tfrac{1}{2}) \quad ,$$

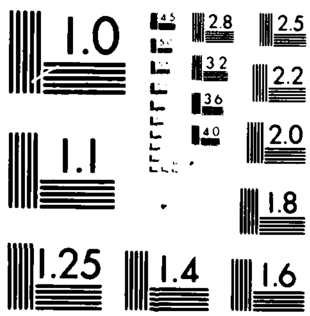
for f_o, the standard normal, where $J_o(u) = \Phi^{-1}(u)$. Hájek and Šidák remark that the Klotz test is asymptotically equivalent to the Capon test where $[J_o(\frac{R_i}{N+1})]^2$ is replaced by its expected value. Our estimators $\hat{\theta}$ and $\hat{\psi}$ are linear transformations of the Van der Waerden and Klotz tests respectively when $\hat{D}(u)$ uses the normal density for f_o. Each is an asympototically optimal test for f_o the standard normal density function.

(ii) The Ansari-Bradley test is

$$T_2 = \sum_{i=1}^{m} [\tfrac{1}{2}(m+n+1) - |R_i - \tfrac{1}{2}(m+n+1)|]$$

AD-A098 240 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
A GENERALIZED APPROACH TO THE TWO SAMPLE PROBLEM: THE QUANTILE --ETC(U)
APR 81 T J PRIHODA DAAG29-80-C-0070
UNCLASSIFIED TR-B-5 ARO-16992.5-M NL

END
DATE
FILMED
5 81
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

$$= \frac{m}{2}(N+1) - \sum_{i=1}^{m}(N+1)\left(\frac{1}{2} - \frac{R_i}{N+1}\right)$$

$$= \frac{m}{2}(N+1) - \frac{m(N+1)}{12}(\hat{\psi}-3),$$

where $\hat{\psi}$ is calculated from Parzen's D(u) with $f_o(x) = \frac{1}{2}(1+|x|)^{-2}$, the density for which the Ansari-Bradley test is also optimal. We note that the Ansari-Bradley test is formed in a manner similar to the Wilcoxon test for location, but with ranks modified to detect the scale difference. Sukhatme (1957) has also introduced a modified Wilcoxon test for scale differences. The Siegel-Tukey test is similar but allows use of the Wilcoxon tables for small samples. Though the *Ansari-Bradley* and Wilcoxon tests are formed similarly they are optimal for different densities.

Through this implementation of Parzen's (1980) approach we obtain both location and scale tests for each density, as well as, the estimates (these were given in Table 1 (p. 9) for completeness). For example, a location difference test and estimate for $f(x) = \frac{1}{2}(1+|x|)^{-2}$ can be obtained from

$$\hat{\theta} = \frac{3}{m}\sum_{i=1}^{m} \text{sign}\left(\frac{1}{2} - \frac{R_i}{N+1}\right) \min\left(\frac{R_i}{N+1}, 1 - \frac{R_i}{N+1}\right).$$

The quartile test for differences in scale seems to be another example of a nonparametric test for one parametric difference which has had no corresponding test for location difference advocated which assumes the same density. In Table 3 (p. 29) we give the test statistics which we advocate for this f_o.

(iii) The Quartile test developed by Westenberg (1948) is

$$T_3 = \frac{1}{2} \sum_{i=1}^{m} [\text{sign}(|R_i - \frac{N+1}{2}| - \frac{N+1}{4}) + 1]$$

$$\simeq \# \text{ of x obs.} \notin (\tilde{H}^{-1}(.25), \tilde{H}^{-1}(.75))$$

$$= m[1 - \tilde{D}(.75) - \tilde{D}(.25)] \ .$$

It is related to the comparison function that we use here and a D(u) model can be obtained from the density for which it is asymptotically optimal, i.e.

$$f(x) = 1 \quad , |x| \leq \frac{1}{4} \ ,$$

$$= \frac{1}{16x^2} \ , |x| > \frac{1}{4} \ .$$

Using this density for f_o in Parzen's D(u) model we obtain

$$\hat{\psi} = \frac{1}{2} - \frac{1}{m} \sum_{R_i \notin [\frac{1}{4}(N+1), \frac{3}{4}(N+1)]} 1$$

for an estimate of the scale differences of the two samples. The location difference estimate obtained simultaneously was given in Table 3 (p. 29).

(iv) The Savage test [I.R. Savage (1956)] is asymptotically optimal for the exponential density and is defined as

$$T = \sum_{i=1}^{m} \sum_{j=N-R_i+1}^{N} j^{-1} \ .$$

The exponential density-quantile function, $f_o[Q_o(u)]$ is not in the RKHS used to obtain $\hat{\psi}$ and $\hat{\theta}$ on the whole interval [0,1]. However, if we truncate the interval to $[\frac{1}{N}, 1 - \frac{1}{N}]$ we may still use all of the data to obtain an estimate of $\hat{D}(u)$. However, we need an algorithm to compute $\hat{\theta}$ and $\hat{\psi}$ for whatever the sample sizes are using the $\langle f_1, f_2 \rangle_{\frac{1}{N}, 1 - \frac{1}{N}}$ as given in Theorem 2.6. This algorithm may also be used to truncate left and/or right portions of the combined sample. The result is quite different from the Savage test and is discussed in section 2.4.

(v) In Table 3 (p. 29) we also gave the scale tests developed from Parzen's $\hat{D}(u)$ model which correspond to the logistic, double exponential, and Cauchy families for f_o. Those $\hat{\psi}$ functions provide formulas for testing equality of scale, and thus extend the set of nonparametric tests at our disposal by combining a location and scale test optimal for the same density.

5.2.2 Exceedance Tests for Scale

A variation of the location Haga test is due to Kamat (1956) and has test statistic T = A + B' - A' - B where these components are defined as in the Haga test (see 5.1.2 (i)). We remark that this exceedance test is also a function of jump sizes in $\tilde{D}_1(u) = \tilde{F}\tilde{G}^{-1}(u)$ and $\tilde{D}_2(u) = \tilde{G}\tilde{F}^{-1}(u)$. Simpler versions of the Kamat test are given by Rosenbaum (1953) and Klotz (1962). We mention this to give an indication of the work done relating to comparison functions other than $\tilde{D}(u) = \tilde{F}\tilde{H}^{-1}(u)$. These tests provide comparisons for further research for Parzen's (1980) $\hat{D}_1(u)$ and $\hat{D}_2(u)$ techniques.

5.2.3 Goodness of Fit Tests for Scale

Hájek and Šidák (1967) remark (p. 99) that one can make the Kolmogorov-Smirnov, Renyi, and Cramér Von Mises tests more sensitive to differences in scale by successively subtracting smallest and largest pairs of C_{D_i} [see Hájek and Šidák (1967)] rather than subtracting $C_{D_1}, \ldots, C_{D_k}$ successively. So, one can compute a goodness of fit test in two ways, one sensitive to location differences and one sensitive to scale differences. Again, we see the attraction and need for simultaneously estimating location and scale differences for a given problem, H_o: $F = G$.

We emphasize that Parzen (1980) has both a location, $\hat{\theta}$, and scale, $\hat{\psi}$, component in the $\hat{D}(u)$ estimator of $D(u) = F[H^{-1}(u)]$, the comparison distribution function, which are asymptotically optimal for the same f_o. In the following section (5.3) we remark on some relationships of Parzen's (1980) methods with various robust, adaptive, and combinations of other techniques.

5.3 Remarks on Some Other Approaches and Extensions

5.3.1 Combinations of Separate Tests for Location and Scale

Duran, Tsai, and Lewis (1976) have combined tests of location and scale to also simultaneously test for equality of both parameters. They use Randles and Hogg's (1971) result, which states that under H_o, even translation invariant statistics (Mood, Klotz, and Ansari-Bradley) are independent of odd translation invariant statistics (Wilcoxon and Normal scores). Then using Chernoff and Savage (1958) they obtain the asymptotic normality of the test statistics under H_o and, with more conditions, an asymptotic bivariate normality result under certain alternatives. Their alternatives are similar to Parzen's D(u) model where we assume θ and ψ small.

They gave no examples, but we still may make some comparisons. More research is needed for their techniques to be evaluated on examples and they did not provide any methods for estimating the location or scale differences. Parzen's approach naturally leads one to simultaneous location and scale tests for the same underlying density which is not the case with the even and odd statistics. For example, combinations could be the Wilcoxon and Ansari-Bradley (different f_o) or Quartile and Median (different f_o) or Normal scores and Klotz (f_o = normal) tests. The analogous result from Parzen's D(u) model is that it is asymptotically optimal for one f_o (examples in Table 1, p. 9). For local alternatives there is a simultaneous test for location and scale. From Corollary 2.2, it is

$$L = N\gamma \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix}' \Sigma \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix} ,$$

and is approximately a $\chi^2(2)$; but, $\hat{\theta}$ and $\hat{\psi}$ also estimate the differences in parameters of the two samples [Parzen (1980)]. In section 3 we use $\hat{D}(u) - \tilde{D}(u)$ to help choose f_o correctly. Further, section 6.3 methods will estimate the differences between samples at any percentile or quantile as well. Lepage (1975, 1976) also gives many results on the distributions and efficiencies of this method of combining tests. Other authors have tried to form tests which are insensitive to differences in one of the parameters while they are sensitive to differences in the other.

5.3.2 Robust and Similar Tests

The classical F-test for $H_o: \frac{\sigma_1}{\sigma_2} = 1$ has been found to be non-robust to deviations from normality with respect to size by many authors. Shorack (1969) examines an approximate permutation test, a "jacknife" procedure, and some "rank like" and other tests for $H_o: \frac{\sigma_1}{\sigma_2} = 1$ by considering their Pitman asymptotic relative efficiency and Monte Carlo studies of power. Shorack's simulation included the uniform, normal, and double exponential densities. The rank like tests do not use all the scale properties of a continuous variable but have other desirable properties as given in Moses (1963). The permutation test (APF-test) is an approximation based on an F statistic and a minor variation (for locations unknown) of an approach used by Box and Anderson (1955). It has the same asymptotic relative efficiency as the F test but a very different robustness level as seen in the Monte Carlo results.

Shorack also inverts this test to obtain a confidence interval for $(\frac{\sigma_1}{\sigma_2})^2$. Another quite practical test which did not do badly in the simulation for $m = n$ was Levene's (1960) test. Although assumptions are violated, Levene suggested doing an ANOVA on the means of $\{(X_i - \bar{X})^2\}$ and $\{(Y_i - \bar{Y})^2\}$. The jacknife like procedure also requires $m = n$ and jacknifes the logarithms of sample variances. It was on a par with the APF test in terms of power. The rank like tests were also an ANOVA of log S_i^2 where S_i is a scale parameter estimator for a subgroup of all n or m observations. The APF test was reported to do quite well in comparison to the others. The approximation of the APF test statistic's distribution makes it attractive for small samples, but it seems quite cumbersome especially in confidence interval calculation. It does not help one choose f_o nor does it decide anything about the location parameter differences.

Some of the earliest attempts to deal with the two sample problem were also approximations. Murphy (1976) compares the t-test, Aspin-Welsh approximate t-test, and Wilcoxon test by simulation. The Aspin-Welsh approximate t statistic is

$$t = (\bar{X} - \bar{Y})/\sqrt{\frac{S_X^{\,2}}{m} + \frac{S_Y^{\,2}}{n}} \quad ,$$

where

$$\frac{1}{df} = \frac{c^2}{(m-1)} + \frac{(1-c)^2}{(n-1)} \quad ,$$

and

$$c = \left(\frac{S_X^2}{m}\right) / \left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right),$$

given by Welsh (1937) with critical values by Aspin (1949). The three tests were compared for normal, uniform, and exponential f_o densities. It is interesting that Murphy concludes the Aspin-Welsh test is highly satisfactory for H_o: $\mu_1 = \mu_2$ when $\sigma_1 \neq \sigma_2$ while the Wilcoxon-Mann-Whitney is not. Murphy also pointed out that no test was satisfactory when skewness was present. We may choose f_o and calculate our results even for f_o skewed. When one assumes f_o normal and $\sigma_1 \neq \sigma_2$, testing H_o: $\mu_1 = \mu_2$ is known as the Behrens-Fisher problem. Sheffé (1970) presents practical solutions to the problem. It appears Behrens (1929) and then Fisher (1935a) began this expansion of the two sample location problem by considering $\sigma_1 \neq \sigma_2$. This was Sir Ronald Fisher's (1935a) controversial paper, "The Fiducial Argument in Statistical Inference." He also proposed nothing less than a randomization test, also in 1935, in his book, The Design of Experiments. Many authors have studied the robustness of the t-test with respect to α. Posten (1978) did an extensive simulation study. He did his study of the t-test over 87 Pearson curve distributions where the level of the test was estimated from 100,000 generated t-values, except for one case (n=30 had "only" 83,000). Posten varied n from 5 to 30, β_1 from 0 to 2, and β_2 from 1.4 to 7.8. Posten points out the obvious conclusions from his tables; i.e., the t is very robust with respect to α when $n = m$.

In fact, all tabulated significance levels round to .04, .05, or .06 through the whole simulation study which had nominal level $\alpha = .05$. Other authors [e.g. Pearson (1931), Geary (1947), Finch (1950), Gayen (1950) and Box (1953)] have shown that this is not the case with the commonly used and taught F-test for variance.

Since the t-test is very robust with respect to α, we should choose a linear rank test based on power or other considerations, not the accuracy of significance levels. In this regard, Fligner and Killeen (1976) have introduced analogues of the Ansari-Bradley, Mood, and Klotz tests which have the same Pitman efficiency, but significantly higher powers for small samples. They respectively are

$$T_1 = [m(n+m)]^{-1} \sum_{i=1}^{m} R_i \quad ,$$

$$T_2 = [m(m+n)^2]^{-1} \sum_{i=1}^{m} R_i^2 \text{ , and}$$

$$T_3 = m^{-1} \sum_{i=1}^{m} \Phi^{-1}[\frac{1}{2} + \frac{R_i}{2(N+1)}]^2 \quad ,$$

where R_i is the rank of $V_i = |X_i - m|$ among the combined sample of V_i's and $W_j = |Y_j - m|$ where m is the median of the combined sample of $\{X_i\}$ and $\{Y_j\}$. These tests may be chosen on the basis of small sample power, where we would choose Parzen's (1980) tests relating to $\hat{D}(u)$ based on simultaneous estimation of θ and ψ or graphical and statistical help in choosing f_o. Perhaps these authors' "score" functions can be interpreted in a way to help develop more small sample estimators for the Parzen approach. Other authors have conjectured, if one can not reduce the influence of nuisance parameters in a particular test, perhaps one can adapt to the influence of the nuisance parameter and obtain a more powerful test of that particular parameter. We, in fact, model both the location and scale differences in the work here.

5.3.3 Adaptive Type Tests

Sen (1962) and Potthoff (1963) have attempted adapting rank tests for location to adjust for unequal variances by a conservative approach. However, many of the rank tests then became dependent on f_o. Others have been more successful. Hogg, Fisher, and Randles (1975) have designed adaptive location test procedures for skewed distributions. Very few nonparametric or robust procedures consider how to detect or what to do when skewness is present. Parzen (1979, 1980) has also given techniques to help detect bimodality, as well as skewness, by using an autoregressive density estimator. Hogg (1976) also remarks on a possible adaptive two sample scale test where one decides to use a Kamat, Klotz, Ansari-Bradley, or quartile test based on the combined order statistics.

5.3.4 Other Approaches of Interest

Korwar and Hollander (1975) have given an empirical Bayes estimator for F(x) which is optimal for a Ferguson Dirichlet prior. Perhaps they would want us to develop a Bayes $\tilde{D}(u)$ and

$\hat{D}(u)$. The first problem is to determine when this prior is adequate.

Censored modifications for the Kolmogorov-Smirnov test have been given by Tsao (1954) and Ishii (1958). Mehrotra and Johnson (1976) extend results in Hájek and Šidák for asymptotically most powerful tests in the two sample problem to apply to censored data, i.e., the first r observations. As mentioned in Parzen (1979, 1980), one can truncate the reproducing kernel Hilbert space estimates by using $\langle f_1, f_2 \rangle_{p,q}$ where $0 < p < q < 1$ or use an inner product based on the censored observations.

Other directions to go include the Wald and Wolfowitz (1940) runs test and any relation it has to these methods. Also, Sen (1963) has investigated a class of tests based on linear combinations of the number of Y_i between $X_{(i)}$ and $X_{(i+1)}$ which can be related to the spacings of the jump points in $\tilde{D}(u)$. Eubank (1979) provides one sample optimal spacings which can be generalized to the two sample problem.

With regard to estimating scale differences, Bhattacharyya (1977) has given techniques based on Sen's (1966) modification for scale parameters of Hodges and Lehmann's technique for estimating location shift. Bhattacharyya provides estimators of σ_1/σ_2 corresponding to the Ansari-Bradley, Siegel-Tukey, and a modified Sukhatme test. Lambscher and Odeh (1976) have also proposed practical methods for estimating scale parameters from a

Sukhatme test. Duran (1976) gives a review of approximately 80 references on tests for scale with many comments on these and other tests. One is the Barton and David (1958) test, not covered here. The great number of techniques Duran comments on makes it impossible to be very detailed for any; but, he gives many valuable comments and references on comparisons of these tests and "minor" modifications of them. Also, Zuijlen (1977) extends much of the rank tests' distribution theory to the non i.i.d. case. Other papers of particular interest deal with comparison function techniques.

5.3.5 Comparison Function Techniques

Wilk and Gnanadesikan (1968) stimulated research in the area of probability plotting where they use Q-Q and P-P plots to compare data sets. A Q-Q plot is essentially a plot of the Y quantile function versus the X quantile function (see section 4.2), the points being joined for a common u, i.e., $G^{-1}F(x)$ versus $X = Q_X(u)$. P-P plots are a plot of $u_X = Q_X^{-1}$ versus $u_Y = Q_Y^{-1}$ where $Q_X = Q_Y$ at each point. Switzer (1976) and Doksum and Sievers (1976) extend the graphical work of Wilk and Gnanadesikan by developing confidence procedures for various comparison functions used in the two sample problem. They estimate a general treatment function, $t(x)$. Since they include the data sets in their papers, we are able to compare their results to those developed here (section 4).

Steck, Zimmer, and Williams (1974) have also developed confidence bands based on $D_2(u) = G[F^{-1}(u)]$ or $D_1(u) = F[G^{-1}(u)]$.

Further research may generalize this to $D(u) = F[H^{-1}(u)]$ and provide some further comparison of the two. Doksum (1974) has *also* given the asymptotic distribution of $t'(x) = \tilde{G}^{-1}[\tilde{F}(x)] - x$. Doksum and Sievers (1976) have also begun developing confidence bands with or without a location scale model being assumed. They also invert two sample statistics for their bands. Their location scale m 'el is

$$t'(x) = \mu_2 + \frac{\sigma_2}{\sigma_1}(x-\mu_1)-x = \mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1 + (\frac{\sigma_2}{\sigma_1} - 1)x$$

$$= \mu_2 - (\psi-1)\ \mu_1 + \psi\ x.$$

In this case, simultaneously estimating $\theta = \frac{\mu_2-\mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$ is not done. However, they have given a likelihood ratio confidence band for f_o normal when $m = n$. Further research could compare this with Parzen's (1980) techniques for estimating

$$D_1'(u) = d(u) = \frac{f[G^{-1}(u)]}{g[G^{-1}(u)]} \quad ,$$

the likelihood ratio for the two samples which does not require $m = n$. Doksum and Sievers show asymptotic equivalence to M.L.E. bands and remark that an advantage of their's is that it can be applied to censored data. Again, Parzen's (1980) techniques have a potential for censored data analysis which can be further explored. They also remark that some of their numerical results show that the general bands are quite inefficient if the correct model is normal. This gives us a motivation to use $\hat{D}(u)$ in helping identify the correct f_o. We do it both for statistical reasons of efficiency and scientific reasons of identifying a correct model. We conclude this section with a few comments on some of the vast amount of research concerning the two sample problem.

5.4 Remarks on the Literature Review

The Behrens-Fisher problem remains open 45 years after the work for which it was named and the list of several more general approximate solutions grows. The approach of Parzen (1980) implemented here has many of the aspects of several of the other authors through the decades. Hopefully, it will contribute to a unified approach by consideration of $\tilde{D}(u)$ and $\hat{D}(u)$ which nearly all the previous techniques are related to in some way. Thus far, only asymptotic properties of $\tilde{D}(u)$ have been given; however, since $\hat{\theta}$ and $\hat{\psi}$ directly relate to linear rank tests, they provide an easy extension to calculation of simultaneous estimates of location and scale differences and use of the finite sample size linear rank tables. This helps unify the techniques of sections 5.1.1 and 5.2.1. We also see the importance of $\tilde{D}(u)$ in the exceedance and goodness of fit tests. The relationship of $\hat{D}(u)$ to $\tilde{D}(u)$ will help utilize goodness of fit tests in choosing the correct linear rank test. By matching $\hat{D}(u)$ to $\tilde{D}(u)$ we will not just adapt the scale differences and estimate the location differences or vice-versa. Rather, we will simultaneously estimate location and scale differences and by comparing the results for various f_o's, determine which family should be assumed for a good fit. This has begun to be explored in sections 3.1 to 3.2.

One method of obtaining a robust estimate is to trim the data. This can be explored by considering $0 < p < q < 1$ rather than $p = 1 - q = 0$.

The recent research on many different comparison function techniques provides techniques to compare with Parzen's (1980) approach using the comparison function, $\tilde{D}(u) = \tilde{F}[\tilde{H}^{-1}(u)]$. This is begun in section 4.

One problem with the two sample research has been simultaneous testing of location and scale differences. This approach clearly will provide a solution, i.e., from Corollary 2.2, we have

$$L = N\gamma \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix}' \sum \begin{pmatrix} \hat{\theta} \\ \hat{\psi} \end{pmatrix} \xrightarrow{D} \chi^2(2) ,$$

under H_o: $F = G$ where $\hat{\theta}$ is the location component and $\hat{\psi}$ is the scale component. One can easily see which difference has a greater change with respect to σ_1 at least. Another problem, especially with nonparametric or distribution free tests, has been to estimate the difference once one has been detected. Parzen's (1980) technique provides local estimators, $\hat{\theta}$ and $\hat{\psi}$, of the difference. With Δ_Q we estimate $\mu_2 - \mu_1$ and $\sigma_2 - \sigma_1$ under H_a. This pair exists for many useful densities and each estimator is asymptotically optimal for a common density. Another problem which the adaptive tests are designed to deal with is to first make a decision about the type of f_o and then an independent test of a parameter difference using the decision about f_o. Parzen's approach using $\tilde{D}(u)$ and $\hat{D}(u)$ gives the asymptotic distribution of $\hat{\theta}$ and $\hat{\psi}$ given f_o. The results in section 3 lead to a minimum distance choice for f_o among the set of f_o that one considers. By examining the $\hat{\theta}$ and $\hat{\psi}$ for each f_o we gain an indication of the importance or lack of importance as to which f_o we should assume. By examining the residuals, $\hat{D}(u) - \tilde{D}(u)$, we use both the location and scale parameters to decide on f_o. However, the estimators of $\hat{\theta}$ and $\hat{\psi}$ are functions of the R_i so a topic of further research could be to try to obtain an independent choice of f_o. Still the scientific interpretations may lead one to consider several models for the data although a statistically more powerful test may exist choosing just one model. In fact, the adaptive answer is to choose f_o based on $\hat{D}$ and then, independently estimate Δ_Q. We may do this, since $\{R_i\}$ are independently distributed of $\{X_{(i)}\}$.

A common problem that robust techniques try to deal with is having a known f_o for the data but shifted location and/or scale for part of the data. As mentioned earlier, using $0 < p < q < 1$ would be a common technique for dealing with this problem. There is an indication the truncation does well with the Cauchy f_o in Rothenberg, Fisher, and Tilanus (1964). This can be further explored. Also one may attempt to model skewness with appropriate f_o in the $D(u)$ model or $\Delta_Q(u)$ model.

Scott, et. al. (1976) have presented a bivariate density estimation technique which also helps deal with bimodality. We also analyze his data set in section 4.3, although not with a bivariate approach. Presently very few approaches attempt to estimate skewness or bimodality differences of two samples. Some indication of how this might be done with a quadratic D(u) model is given in section 6. For other directions of research see Parzen (1979, 1980) or section 6.

6. SOME ALTERNATIVE MODELS FOR D(u)

In our approach, just as in ordinary regression, one will often consider more than one model for the data.

With our D(u) model we assume an f_o family for the underlying density, although we give techniques to help choose it. We also assume a linear Taylor series expansion is adequate and that θ and ψ are small. If θ and ψ are found to be of moderate size we may wish to improve the expansion by including more terms as in Section 6.1.

Rather than including more terms, we may use the Δ_Q (see 1.10) model which is accurate under the alternatives $\theta \neq 0$ and/or $\psi \neq 0$. In section 6.2 we do this, still assuming that the underlying family is the same for both the X_i and Y_j. There we suggest estimators of $\mu_2 - \mu_1$ rather than $\theta = \frac{\mu_2-\mu_1}{\sigma_1}$ and $\sigma_2 - \sigma_1$ rather than $\psi = \frac{\sigma_2-\sigma_1}{\sigma_1}$. If we had no idea which f_o would model the data, we may wish to construct a model which converges to the correct f_o and is, in a sense, f_o free. In the past the convergence has been quite slow and tests have been less powerful than those which assume f_o known. However, in section 6.3 we suggest methods to be explored which make still fewer assumptions than those made thus far regarding f_o.

6.1 The Quadratic Model

In this section we give a quadratic expansion which results in an alternate model of D(u) containing some of the quadratic terms, i.e. those of the quadratic expansion of F_o about $\frac{x-\mu_1}{\sigma_1}$.

In section 2 we represent $G(x)$ in terms of F_o as follows (when θ and ψ are small)

$$G(x) \doteq F_o[(\frac{x-\mu_1}{\sigma_1}) - \theta - \psi(\frac{x-\mu_1}{\sigma_1})]$$

A quadratic Taylor series expansion of $F_o(\frac{x-\mu_1}{\sigma_1})$ about $\frac{x-\mu_1}{\sigma_1}$ gives

$$G(x) \doteq F_o(\frac{x-\mu_1}{\sigma_1}) - f_o(\frac{x-\mu_1}{\sigma_1})[\theta+\psi(\frac{x-\mu_1}{\sigma_1})]$$

$$+ \tfrac{1}{2} f_o'(\frac{x-\mu_1}{\sigma_1})[\theta+\psi(\frac{x-\mu_1}{\sigma_1})]^2$$

Again, letting $x = H^{-1}(u) \doteq F^{-1}(u) = \mu_1+\sigma_1 Q_o(u)$ as on p. 18, gives $Q_o(u) \doteq \frac{x-\mu_1}{\sigma_1}$ and

$$G[H^{-1}(u)] \doteq FH^{-1}(u) - f_oQ_o(u)[\theta+\psi Q_o(u)] + \tfrac{1}{2} f_oQ_o(u)J_o(u)[\theta^2+2\theta\psi Q_o(u)$$

$$+ \psi^2 Q_o^2(u)] .$$

Since $H(x) = \lambda F(x) + (1-\lambda)G(x)$, we have for $x = H^{-1}(u)$

$$HH^{-1}(u) \doteq \lambda FH^{-1}(u) + (1-\lambda)[FH^{-1}(u) - \theta f_oQ_o(u) - \psi Q_o(u) f_oQ_o(u) + \frac{\theta^2}{2} f_oQ_o(u)J_o(u)$$

$$+ \theta\psi f_oQ_o(u)J_o(u)Q_o(u) + \frac{\psi^2}{2} f_oQ_o(u)J_o(u)Q_o^2(u)] ,$$

and, since $D(u) = FH^{-1}(u)$, we have

$$u \doteq D(u) - (1-\lambda)[\theta f_oQ_o(u) + \psi Q_o(u) f_oQ_o(u) + \theta' f_oQ_o(u)J_o(u)$$

$$+ \gamma' f_oQ_o(u)J_o(u)Q_o(u) + \psi' f_oQ_o(u)J_o(u)Q_o^2(u)]$$

where $\theta' = -\frac{1}{2}\theta^2$, $\gamma' = -\frac{1}{2}\theta\psi$, and $\psi' = -\frac{1}{2}\psi^2$. Finally,

$$D(u)-u \doteq (1-\lambda)[\theta f_1(u)+\psi f_2(u)+\theta' f_3(u)+\gamma' f_4(u)+\psi' f_5(u)],$$

where $f_1(u) = f_oQ_o(u), f_2(u) = Q_o(u)f_oQ_o(u), f_3(u) = J_o(u)f_oQ_o(u)$, $f_4(u) = Q_o(u)J_o(u)f_oQ_o(u)$, and $f_5(u) = Q_o^2(u)J_o(u)f_oQ_o(u)$.

If the $f_i(u)$ are in the RKHS of $B(u)$ with $p = 1 - q = 0$, then we may estimate θ, ψ, θ', γ', and ψ' provided we treat θ', γ', and ψ' as free parameters. For computational convenience we would do this to begin with. This model gives us a 5 x 5 matrix, Σ_5, rather than a 2 x 2 Σ as in the linear expansion. In fact, for Σ_5 and g_5 in this quadratic expansion we need fifteen and five inner products to exist respectively. One may try to orthogonalize Σ_5 to reduce the problem. We suspect the shapes of $f_oQ_o(u)J_o(u) \cdot Q_o(u)$ and $f_oQ_o(u)J_o(u)Q_o^2(u)$ may be useful detectors of bimodality and skewness respectively. We leave this for further research.

6.2 The Δ_Q Model

As shown in Theorem 1.1, Parzen's (1979) model for Q(u) in the one sample and scale problem gives an asymptotically exact two sample model for $\Delta_Q(u)$ also. In this section we further develop this model by suggesting estimators of $\mu_2-\mu_1$, $\sigma_2-\sigma_1$, and $\Delta_Q(u)$ based on Parzen (1961, 1967). We also give the asymptotic distribution of these estimators and some remarks on their use in the analysis of two sample data.

The model for $\Delta_Q(u)$ suggested by Theorem 1.1 is

$$f_oQ_o(u)\Delta_Q(u)=\Delta_\mu f_oQ_o(u)+\Delta_\sigma Q_o(u)f_oQ_o(u) \ ,$$

with the following estimator

$$f_oQ_o(u)\hat{\Delta}_Q(u)=\hat{\Delta}_\mu f_oQ_o(u)+\hat{\Delta}_\sigma Q_o(u)f_oQ_o(u)$$

obtained using Parzen (1979) results.

We suggest estimators of $\Delta_\mu = \mu_2 - \mu_1$ and $\Delta_\sigma = \sigma_2 - \sigma_1$ in Theorem 6.1 and give their asymptotic distribution.

Theorem 6.1: If the conditions of Theorem 1.1 hold and f_oQ_o and $Q_o(f_oQ_o)$ are members of the RKHS of B(u) for p = 1 - q = 0, then as $N \to \infty$ such that $\lambda_N = \frac{m}{N} \to \lambda_o$ $(0 < \lambda_o < 1)$, we have

$$\sqrt{N}\begin{bmatrix}\hat{\Delta}_\mu - \Delta_\mu \\ \hat{\Delta}_\sigma - \Delta_\sigma\end{bmatrix} \xrightarrow{D} N_2\left[\binom{0}{0},\ c_2{}^2\Sigma^{-1}\right]$$

where $\hat{\Delta}_\mu = \hat{\mu}_2-\hat{\mu}_1$ $\hat{\Delta}_\sigma = \hat{\sigma}_2-\hat{\sigma}_1$, $c_2{}^2 = \lambda_o\sigma_1{}^2+(1-\lambda_o)\sigma_2{}^2$, and $\hat{\mu}_i$ and $\hat{\sigma}_i$ are as given in Parzen (1979).

Proof: From Csörgö and Révész (1978), Parzen (1979), and Eubank (1979) we obtain for i = 1 or 2, where n_i denotes the i^{th} sample size,

$$\sqrt{n_i}\ f_oQ_o(u)[\tilde{Q}_i(u)-\mu_i-\sigma_iQ_o(u)] \xrightarrow{L} \sigma_i\ B(u) \ ,$$

and further [using Parzen (1961, 1967)],

$$\sqrt{n}_i \begin{pmatrix} \hat{\mu}_i - \mu_i \\ \hat{\sigma}_i - \sigma_i \end{pmatrix} \xrightarrow{D} N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_i^2 \Sigma^{-1} \right].$$

Letting $\hat{\Delta}_\mu = \hat{\mu}_2 - \hat{\mu}_1$ and $\hat{\Delta}_\sigma = \hat{\sigma}_2 - \hat{\sigma}_1$, we obtain

$$\sqrt{N} \begin{pmatrix} \hat{\Delta}_\mu - \Delta_\mu \\ \hat{\Delta}_\sigma - \Delta_\sigma \end{pmatrix} \xrightarrow{D} N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, c_2^2 \Sigma^{-1} \right]$$

where $c_2^2 = \lambda_o \sigma_1^2 + (1-\lambda_o)\sigma_2^2$, since $N = n_1 \lambda_N^{-1} = n_2(1-\lambda_N)^{-1}$ and $\lambda_N \to \lambda_o$ as $N \to \infty$, and linear combinations of independently distributed random variables converge to the linear combination of their asymptotic limits.

<u>Corollary 6.1</u>: If the conditions of Theorem 6.1 hold, then $\hat{\Delta}_\mu$ and $\hat{\Delta}_\sigma$ are given by

$$\begin{pmatrix} \hat{\Delta}_\mu \\ \hat{\Delta}_\sigma \end{pmatrix} = \Sigma^{-1} \mathbf{g}_2 - \Sigma^{-1} \mathbf{g}_1 = \Sigma^{-1} \mathbf{g}_3 ,$$

where Σ is given in section 2, $\mathbf{g}_1$ and $\mathbf{g}_2$ are given in Parzen (1979) and Eubank (1979) for i = 1, 2 as

$$\mathbf{g}_i = \begin{bmatrix} \langle f_o Q_o, (f_o Q_o)\tilde{Q}_i \rangle \\ \langle Q_o(f_o Q_o), (f_o Q_o)\tilde{Q}_i \rangle \end{bmatrix} ,$$

and we define $\mathbf{g}_3 = \mathbf{g}_2 - \mathbf{g}_1$.

<u>Proof:</u> By definition of $\hat{\Delta}_\mu$ and $\hat{\Delta}_\sigma$, $\begin{bmatrix} \hat{\Delta}_\mu \\ \hat{\Delta}_\sigma \end{bmatrix} = \Sigma^{-1}\mathbf{g}_2 - \Sigma^{-1}\mathbf{g}_1$.

By definition of matrix operations, $\Sigma^{-1}\mathbf{g}_2 - \Sigma^{-1}\mathbf{g}_1 = \Sigma^{-1}(\mathbf{g}_2 - \mathbf{g}_1) = \Sigma^{-1}\mathbf{g}_3$, since $\mathbf{g}_3 = \mathbf{g}_2 - \mathbf{g}_1$. Then, by definition of inner products, since

$\underline{g}_3 = \underline{g}_2 - \underline{g}_1$, we note that

$$\underline{g}_3 = \begin{bmatrix} \langle f_oQ_o, f_oQ_o(\tilde{Q}_2-\tilde{Q}_1)\rangle \\ \langle Q_o(f_oQ_o), f_oQ_o(\tilde{Q}_2-\tilde{Q}_1)\rangle \end{bmatrix}$$

<u>Remarks</u>: We call $\tilde{\Delta}_Q(u) = \tilde{Q}_Y(u) - \tilde{Q}_X(u)$ the raw difference of quantile functions at the quantile u and $\hat{\Delta}_Q(u) = \hat{Q}_Y(u) - \hat{Q}_X(u)$ the estimated difference of quantile functions at the quantile u. These names are suggestive of our interpretation of $\Delta_Q(u)$. Note that this interpretation and model of $\Delta_Q(u)$ are asymptotically exact under <u>all</u> location and scale alternatives of H_o: F = G, i.e., $\theta \neq 0$ and $\psi \neq 0$. However, since c_2 involves the scale parameters, we note that in using the estimators suggested here, as in Parzen (1979) and Eubank (1979), we presently need to treat c_2 as a free parameter. The implementation and adequacy of the treatment and model is a problem for further research. We emphasize that $\Delta_Q(u)$ may be estimated independently of D(u).

Next we give a definition and the asymptotic distribution of $\hat{\Delta}_Q(u)$, $0 < u < 1$. Let $\hat{\Delta}_s(u) = \sqrt{N}\, f_oQ_o(u)[\hat{\Delta}_Q(u) - \Delta_Q(u)]$ be the standardized $\hat{\Delta}_Q(u)$ so that $\Delta_s(u) = 0$ for $0 < u < 1$.

<u>Theorem 6.2</u>: If the conditions of Theorem 6.1 hold and f_o is symmetric, for $\{u_i \in (0,1);\ i = 1, \ldots, k\}$ and $\underline{\hat{\Delta}}_s(\underline{u}) = [\hat{\Delta}_s(u_1), \hat{\Delta}_s(u_2), \ldots, \hat{\Delta}_s(u_k)]'$, then as $N \to \infty$ such that $\lambda_N = \frac{m}{N} \to \lambda_o$ $(0 < \lambda_o < 1)$, we have

$$\underline{\hat{\Delta}}_s(\underline{u}) \xrightarrow{D} N_k(\underline{0}, c_2^{\,2}\, \Sigma_k)$$

where $\Sigma_k = (\sigma_{ij})$ (note Σ_k is singular when $k > 2$) and

$$\sigma_{ij} = \frac{f_oQ_o(u_i)f_oQ_o(u_j)}{\int_0^1 J_o^{\,2}(u)\,du} + \frac{Q_o(u_i)f_oQ_o(u_j)f_oQ_o(u_j)}{\int_0^1 [1-Q_o(u)J_o(u)]^2\,du} \quad .$$

Proof: Since

$$\hat{\Delta}_s(u) = L\begin{pmatrix}\hat{\Delta}_\mu \\ \hat{\Delta}_\sigma\end{pmatrix} \quad ,$$

where

$$L = \sqrt{N}\begin{bmatrix} f_oQ_o(u_1) & Q_o(u_1)f_oQ_o(u_1) \\ f_oQ_o(u_2) & Q_o(u_2)f_oQ_o(u_2) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ f_oQ_o(u_k) & Q_o(u_k)f_oQ_o(u_k) \end{bmatrix}$$

and since, $\begin{pmatrix}\hat{\Delta}_\mu \\ \hat{\Delta}_\sigma\end{pmatrix} \xrightarrow{D} N_2(\underline{0}, c_2^{\,2}\Sigma^{-1})$, we have

$$\underline{\hat{\Delta}}_s(\underline{u}) \xrightarrow{D} N_k(\underline{0},\ L\,c_2^{\,2}\,\Sigma^{-1}\,L') \quad .$$

Clearly, $\Sigma_k = L\,\Sigma^{-1}L' = (\sigma_{ij})$ implies σ_{ij} is as desired.

Remarks: For this $\hat{\Delta}_Q(u)$ estimator we may directly apply the results of Eubank (1979) and choose $\{u_i;\ i = 1, \ldots, k\}$ for small k as he suggests. We may also use $\{u_i;\ i = 1, \ldots, N\}$ where $\tilde{Q}_Y(u)$ or $\tilde{Q}_X(u)$ have jump points, i.e., u_i corresponding to the data points.

A model for $\Delta_Q(u)$ provides many problems for further research besides those mentioned thus far. For example, as Professor W. C. Parr has pointed out to me, if a plot of $\tilde{\Delta}_Q(u)$ versus $Q_o(u)$ is linear, then $\tilde{F}$ and $\tilde{G}$ are location and scale shifts of the distribution corresponding to $Q_o(u)$. Further, the intercept of the vertical axis is $\mu_2 - \mu_1$ and the slope of the line is $\sigma_2 - \sigma_1$. Tests and estimates based on this fact are a topic of further research. We also leave the Brownian bridge representation of $\hat{Q}(u)$, $\hat{\Delta}_Q(u)$, and $\hat{\Delta}_Q(u) - \tilde{\Delta}_Q(u)$ as topics for further research. We remark that the residuals, $\hat{\Delta}_Q(u) - \tilde{\Delta}_Q(u)$, once their distribution was derived, could be used to select an appropriate f_o to model the data for any location and scale alternative hypothesis of H_o: $F = G$ for independent samples.

Finally, although we do not address the k-sample problem in this work we offer Theorem 6.3 for the following definition of the k-sample problem to suggest further research.

Suppose we have $k \geq 2$ independent random samples, denoted by

$$\{X_{ji};\ j = 1, \ldots, k\ ;\ i = 1, \ldots, n_i\}$$

where n_j is the sample size of the jth random sample and each sample is n_j realizations of the j^{th} random variable X_j. Further, suppose X_j has distribution function $F_j(x) = F_o(\frac{x-\mu_j}{\sigma_j})$ where F_o satisfies the conditions of Theorem 6.1. This is essentially a generalized analysis of variance problem studied by White (1981) and similar to a problem studied by Hájek and Šidák (1967), chapter 3, section 4, and Sen (1962), but more general. In the following theorem we may study a general contrast of the location parameters and the scale parameters simultaneously for the k populations.

Theorem 6.3: For this definition of the k-sample problem, let $\{a_j;\ j = 1, \ldots, k\}$ be fixed constants, $\ell_Q = \sum_{j=1}^{k} a_j Q_j(u)$, $\ell_\mu = \sum_{j=1}^{k} a_j \mu_j$, and $\ell_\sigma = \sum_{j=1}^{k} a_j \sigma_j$. Then as $N = \sum_{j=1}^{k} n_j \to \infty$ such that $\lambda_{Nj} = \frac{n_j}{N} \to \lambda_{oj}$ $(0 < \lambda_{oj} < 1)$; $j = 1, \ldots, k$, we have

$$\sqrt{N}\ f_o Q_o(u)[\tilde{\ell}_Q(u) - \ell_\mu - \ell_\sigma\ Q_o(u)] \xrightarrow{L} c_3\ B(u)$$

and

$$\sqrt{N} \begin{pmatrix} \hat{\ell}_\mu - \ell_\mu \\ \hat{\ell}_\sigma - \ell_\sigma \end{pmatrix} \xrightarrow{D} N_2[\binom{0}{0},\ c_3^2 \Sigma^{-1}],$$

where $\hat{\ell}_\mu = \sum_{j=1}^{k} a_j \hat{\mu}_j$ and $\hat{\ell}_\sigma = \sum_{j=1}^{k} a_j\ \hat{\sigma}_j$, and $\hat{\mu}_j$ and $\hat{\sigma}_j$ are as in Parzen (1979) and Eubank (1979). Finally, $c_3^2 = \sum_{j=1}^{k} a_j^2 \lambda_{oj}\ \sigma_j^2$.

<u>Proof:</u> $\sum_{j=1}^{k} a_j \lambda_{oj}^{1/2} \sigma_j B_j(u) = c_3 B(u)$ since all the $B_j(u)$ are independent Brownian bridges. This gives the first result. Similarly,

$$\sqrt{N} \sum_{j=1}^{k} a_j \begin{pmatrix} \hat{\mu}_j - \mu_j \\ \hat{\sigma}_j - \sigma_j \end{pmatrix} = \sqrt{N} \begin{pmatrix} \hat{\ell}_\mu - \ell_\mu \\ \hat{\ell}_\sigma - \ell_\sigma \end{pmatrix}$$

gives

$$\left(\sum_{j=1}^{k} a_j^2 \lambda_{oj} \sigma_j^2\right) \Sigma^{-1} = c_3^2\, \Sigma^{-1},$$

for the variance-covariance matrix needed. Clearly, the asymptotic mean is zero.

6.3 Raw $\tilde{Q}$ and $\tilde{J}$ Estimators

For the D(u) model suggested by Parzen (1980) we have implemented estimators of $\theta = \frac{\mu_2 - \mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2 - \sigma_1}{\sigma_1}$ given f_o. Our $\Delta_Q(u)$ model also depends on f_o. A topic of further research is to develop estimators which converge to f_o and provide estimators of the location and scale parameters. In this section we suggest an approach to this topic.

Suppose we accept the linear approximation in $\hat{D}(u)$ but do not have a viable choice for f_o. We may then consider studying the technique proposed in this section. For the inner products in $\Sigma^{-1}\,\underline{g}$ we only need Q_o and J_o. Parzen (1979) gives a consistent estimate of Q_o in $\tilde{Q}$ and Hájek and Šidák (1967, p. 260 equation (7)) give a consistent estimate of J, denoted $\tilde{J}$, which are each functions of the order statistics.

Then Parzen's (1980) model yields (for symmetric f_o)

$$\hat{\theta} = \frac{\int_0^1 [-J_o(u)]d[\tilde{D}(u)-u]}{\int_0^1 [J_o(u)]^2 \, du}$$

and

$$\hat{\psi} = \frac{\int_0^1 [1-Q_o(u)J_o(u)]d[\tilde{D}(u)-u]}{\int_0^1 [1-Q_o(u)J_o(u)]^2 du} ,$$

which are solely functions of Q_o, J_o, and the data, $\tilde{D}$. We need appropriate definitions of J and Q using the data based on $\tilde{Q}$ and $\tilde{J}$ perhaps. Further research may explore these estimators from Parzen (1979) and Hájek and Šidák (1967).

One of the difficulties in the problem would be to combine the two samples' different $\tilde{Q}$ and $\tilde{J}$ to obtain the θ and ψ estimators.

7. EVALUATION OF D(u) THROUGH SIMULATION EXAMPLES

As in any regression model, our regression model for D(u) from Parzen (1980) may not contain the correct independent variables and error term. We have added sections 2, 3, 4, and 5 to Parzen's arguments for using the D(u) model and show how it provides useful information whether we reject $H_o: F = G$ or not. As suggested in the remarks of sections 4 and 5 and in the confidence regions for θ and/or ψ in section 2, we desire to make inferences using $\hat{\theta}$, $\hat{\psi}$, and $\hat{D}(u)$ when we detect that $\theta \neq 0$ or $\psi \neq 0$. We also provide the $\Delta_Q(u)$ model which we know may be used when $\theta \neq 0$ or $\psi \neq 0$. As mentioned in sections 2.4 and 4.2, we also become interested in trimming our estimates of θ and ψ for some particular densities or in the presence of suspected outliers. In section 7.1 we make remarks on a design for a simulation study to evaluate the accuracy of the D(u) estimator. In section 7.2 we give a few simulated examples with "large" θ and ψ for six different densities and $m = n = 30$.

7.1 Remarks on Factors of Interest in a Simulation Study

In this section we propose that a simulation study of D(u) and $\Delta_Q(u)$ include:

(1) investigation of the effects of θ, ψ, n, and m on the estimates in $\hat{D}(u)$ and $\hat{\Delta}_Q(u)$ for various f_o, and

(2) investigation of the potential value of truncation of the estimators in $\hat{D}(u)$ and $\hat{\Delta}_Q(u)$ when the two samples have some contaminated observations.

Each of the factors involved in a design should be at several levels. We propose θ, ψ, n, m, f_o, and contamination as the factors. Some dependent variables of interest are:

(1) $\hat{D}(u_i) - D(u_i)$, $\hat{\Delta}_Q(u_i) - \Delta_Q(u_i)$,

(2) $\tilde{D}(u_i) - D(u_i)$, $\tilde{\Delta}_Q(u_i) - \Delta_Q(u_i)$,

(3) $\hat{\theta} - \theta$, $\hat{\Delta}_\mu - \Delta_\mu$,

(4) $\hat{\psi} - \psi$, $\hat{\Delta}_\sigma - \Delta_\sigma$,

and various functions of these quantities, for example, mean square error, bias, and variance of the estimates. We would expect the main effect of each factor to be significant in predicting most of the dependent variables. Also, if the $\theta \times \psi$ interaction were not significant in its effect on a dependent variable, as we hope for small and moderate θ and ψ, then the simultaneous estimation of location and scale parameter differences will pose no problem beyond the ordinary estimation problems of an individual location or scale difference that researchers have traditionally dealt with. The implementation of this simulation study is a topic of further research. The next section reports on six simulated examples.

7.2 Simulated Examples

While in section 4 we compare the approach here with analysis of "live" data sets from other research, in this section we report on a few simulated examples to begin to explore the situations where our D(u) model will obtain reliable results. These examples demonstrate the need for further research and understanding of the techniques developed in this work.

We generated six data sets and submitted them to analysis. All six pairs of samples used $m = n = 30$, $\theta = .5$, and $\psi = .5$. One pair was generated from each of the six distributions given in Tables 5a and 5b. This means each of the $\hat{\theta}$ in Table 5a and each of the $\hat{\psi}$ in Table 5b are estimating the true value of .5. The N, L, C, D.E., A.B., and Q denote the normal, logistic, Cauchy, double exponential, "Ansari-Bradley", and "quartile" densities respectively. The '*' by an estimate denotes the estimate is beyond two standard deviations (under H_o: $F = G$) from its true value of .5.

The theoretical error rate under H_o: $\theta = \psi = 0$ is approximately $.05 + .05 - (.05)^2 = .0975$ for a given f_o. Since we have no replications we are unable to draw any conclusions.

It is pleasing that 56 of the 72 nonparametric estimates were within two standard deviations of the true θ or ψ for two reasons, although we make no conclusions regarding the results without replications. One reason is that five of the six columns of Table 5 have estimates from a nonoptimal f_o. The other reason is that $\sigma_{\hat{\theta}}$ and $\sigma_{\hat{\psi}}$ are derived under H_o: $F = G$, rather than nonzero values of θ and ψ

5a. Simulated $\hat{\theta}$ Fxamples

$\sigma_{\hat{\theta}}$	$\hat{\theta}$ Assumed f_o	True f_o N	L	C	D.E.	A.B.	Q
.26	N	.22	.50	.62	.21	.20	.86
.45	L	.31	.87	1.15	.47	.31	1.72*
.37	C	-.08	.27	.82	.44	.20	1.41*
.26	D.E.	0	.27	.53	.4	.13	.93
.22	A.B.	-.16*	-.03*	.23	.37	.04*	.54
.08	Q	.01*	.16*	.08*	.04*	.007*	.23*

5b. Simulated $\hat{\psi}$ Examples

$\sigma_{\hat{\psi}}$	$\hat{\psi}$ Assumed f_o	True f_o N	L	C	D.E.	A.B.	Q
.18	N	.52	.29	.24	.26	.40	.37
.22	L	.68	.35	.30	.31	.50	.44
.16	C	.29	.12*	.12*	.08*	.16*	.12*
.26	D.F.	.83	.43	.39	.37	.60	.52
.63	A.B.	1.74	.77	.80	.50	1.07	.75
.26	Q	.93	.27	.27	.13	.27	.27

$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$ and $\sigma_{\hat{\psi}} = \sqrt{V(\hat{\psi})}$ under H_o: F=G as in these examples. Perhaps some questions of interest suggested by Tables 5a and 5b would involve testing location parameters with f_o the "quartile" density and testing scale parameters with f_o the Cauchy density.

A definite topic of further research is to determine what values of θ and ψ may be used with reliable results for the D(u) model.

8. CONCLUDING REMARKS

Here we summarize what we have done, what we have not done, and make suggestions for further investigation and implementation. We begin with the mathematical results.

8.1 The Mathematical Problem

With Parzen's (1961, 1979, 1980) time series regression models using the quantile function we have provided new theory and methods for studying how two samples differ in location and scale parameters and at all quantiles. These methods assume continuous increasing F_o and all but the exponential densities were symmetric about zero. Nevertheless, the body of simple linear rank theory and methods for location and scale parameter differences has been expanded and made more complete. The test obtained using Parzen's (1980) D(u) model is a simultaneous location and scale test when the two population distributions are a location and scale shift of a common distribution. These tests are nonparametric, but still provide estimators of the location differences by $\hat{\theta}$ and the scale differences by $\hat{\psi}$ when f_o is the correct density. We also give computational formulas for $\hat{\theta}$ and $\hat{\psi}$ simultaneously or individually for several underlying densities.

By examining the residuals, $\hat{D}(u) - \tilde{D}(u)$, for a finite set of u values we are given some guidance in selecting the underlying density f_o which seems to model the data better than others. This also provides a criteria for selecting which set of nonparametric tests and estimators to use.

There is always the possibility that $\hat{D}(u)$ will not fit $\tilde{D}(u)$ well. By examining $\hat{D}(u) - \tilde{D}(u)$ at different values of u we may see which quantiles contribute more to the deviation of $\hat{D}(u)$ from $\tilde{D}(u)$. The willing user may also suggest his own f_o and go through the estimation and testing calculations to obtain another f_o which may model the data better. In any event, our significance levels are correct as given in Corollary 2.2 and Theorem 2.7. The other confusing possibility would be that several f_o would fit the data well. In this case we would need to check if all $\hat{\theta}$ and $\hat{\psi}$ were consistent and remark that a larger sample size will be more discriminating. Eubank's (1979) optimal u_i for a given density may become valuable in this discrimination process. Also of interest here are the alternate models of D(u).

We discuss what consequence these techniques have for the scientist who analyzes two sample data.

8.2 The Scientific Problem

Given two samples from an experiment, often a treatment and control group, the scientist is faced with determining how the two samples differ and trying to model and explain that difference. This implementation of Parzen's (1980) models provides a general location and scale approach. The test statistics and estimators for parametric differences are easily calculated from the ranks of the X observations in the combined sample of X's and Y's and f_o. Where an existing simple linear rank statistic is a linear transform of $\hat{\theta}$ or $\hat{\psi}$, there are finite sample size tables which may be used if n and m are not large.

There are also several graphical comparisons of the two samples provided. The slope of $\hat{D}(u)$ provides a likelihood ratio type comparison function of the two samples at each quantile. We also provide a graphical comparison of the differences of the two samples at each quantile, i.e., $\tilde{Q}_Y - \tilde{Q}_X$.

Besides the graphical comparisons for Δ_Q one may develop tests of H_o: $F = G$ versus a difference in location and scale. Also provided are the tests and estimates of location and scale difference by considering $\theta = \frac{\mu_2 - \mu_1}{\sigma_1}$ and $\psi = \frac{\sigma_2 - \sigma_1}{\sigma_1}$. In addition, whether these differences are zero or not, $\hat{D}(u) - \tilde{D}(u)$ provides a criteria for choosing an adequate density to model the data. One may also leave one of the differences out and see whether the other difference alone is an adequate model of the differences of the two samples. That is, one may use $\hat{D}(u) - u = (1-\lambda)\ \hat{\theta}\ f_oQ_o(u)$ or $\hat{D}(u) - u = (1-\lambda)\ \hat{\psi}\ Q_o(u)\ f_oQ_o(u)$ rather than both terms at once, as illustrated with the exponential in Theorem 2.3. In this case, we merely drop some terms from the distributions developed for $\hat{D}(u)$, $\hat{D}(u) - \tilde{D}(u)$, and $\hat{\Delta}_Q(u)$.

REFERENCES

Aspin, Alice A. (1949), "Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated," Biometrika, 36, 290-292.

Barton, D. E. and David, F. N. (1958), "A Test for Birth Order Effects," Annals of Eugenics, 22, 250-257.

Behrans, W. V. (1929), "Ein Betrag zur Fehlenberecherung bei wenigen Beobachtungen," Landw. Jb., 68, 807-837.

Bhattacharyya, Helen T. (1977), "Nonparametric Estimation of Ratio of Scale Parameters," Journal of the American Statistical Association, 358, 459-463.

Box, G. E. P. (1953), "Non-Normality and Tests on Variances," Biometrika, 40, 318-335.

Box, G. E. P. and Anderson, S. L. (1955), "Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumptions," Journal of the Royal Statistical Society, Series B, 17, 1-26.

Brown, Mark (1970), "Convergence in Distribution of Stochastic Integrals," Annals of Mathematical Statistics, 41, 829-842.

Chernoff, Herman and Savage, I. Richard (1958), "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," Annals of Mathematical Statistics, 29, 972-994.

Csörgö, M. and Révész, P. (1978), "Strong Approximations of the Quantile Process," Annals of Statistics, 6, 882-894.

Doksum, Kjell A. (1974), "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two Sample Case," Annals of Statistics, 2, 267-77.

Doksum, Kjell A. and Sievers, Gerald L. (1976), "Plotting with Confidence: Graphical Comparison of Two Populations," Biometrika, 63, 421-434.

Duran, Benjamin S. (1976), "A Survey of Nonparametric Tests for Scale," Communications in Statistics: Theory and Methods, A5(14), 1287-1312.

Duran, Benjamin S., Tsai, W. and Lewis, T. (1976), "A Class of Location and Scale Nonparametric Tests," Biometrika, 63, 173-176.

Durbin, J. (1973), Distribution Theory for Tests Based on the Sample Distribution Function, Regional Conference Series in Applied Mathematics, 9, Philadelphia, Society for Industrial and Applied Mathematics.

Eubank, Randall L. (1979), "A Density-Quantile Function Approach to Choosing Order Statistics for the Estimation of Location and Scale Parameters," Texas A&M University, Institute of Statistics Tech. Report No. A-10, 1-119, College Station, Texas.

Finch, D. J. (1950), "The Effect of Non-Normality on the Z-Test When Used to Compare the Variances in Two Populations," Biometrika, 37, 187-189.

Fisher, Ronald A. (1935a), The Design of Experiments, Edinburgh, Oliver and Boyd.

Fisher, Ronald A. (1935b), "The Fiducial Argument in Statistical Inference, Annals of Eugenics, 6, 391-398.

Fligner, Michael A. and Killeen, Timothy J. (1976), "Distribution-Free Two Sample Tests for Scale," Journal of the American Statistical Association, 71, 210-213.

Gayen, A. K. (1950), "The Distribution of the Variance Ratio in Random Samples of Any Size Drawn from Non-Normal Universes," Biometrika, 37, 236-255.

Geary, R. C. (1947), "Testing for Normality," Biometrika, 37, 209-241.

Haga, T. (1959), "A Two Sample Rank Test on Location," Annals of the Institute of Statistical Mathematics, 11, 211-219.

Hájek, Jaroslav and Šidák, Zbyněk (1967), Theory of Rank Tests, New York, Academic Press.

Hodges, J. L. Jr. and Lehmann, E. L. (1963), "Estimation of Location Based on Rank Tests," Annals of Mathematical Statistics, 34, 598-611.

Hogg, Robert V. (1976), "A New Dimension to Nonparametric Tests," Communications in Statistics: Theory and Methods, A5(14), 1313-1325.

Hogg, Robert V., Fisher, D.M. and Randles, R. H. (1975), "A Two Sample Adaptive Distribution-Free Test," Journal of the American Statistical Association, 70, 656-661.

Ishii, G. (1958), "Kolmogorov-Smirnov Test in Life Test," Annals of the Institute of Statistical Mathematics, 10, 37-46.

Kamat, A. R. (1956), "A Two Sample Distribution-Free Test," Biometrika, 43, 377-387.

Klotz, J. (1962), "Nonparametric Tests for Scale," Annals of Mathematical Statistics, 33, 498-512.

Kolmogorov, A. N. (1933), "Sulla Determinazione Empirica di Una Legge di Distribuzione, Giorn. dell'Instituto Ital. Degli Attuari, 4, 83-91.

Korwar, Ramesh M. and Hollander, Myles (1975), "Empirical Bayes Estimation of a Distribution Function," Annals of Statistics, 4, 581-588.

Kshirsagar, Anant (1972), Multivariate Analysis, New York, Marcel Dekker.

Lambsher, N. F. and Odeh, R. E. (1976), "A Confidence Interval for the Scale Parameter Based on Sukhatme's Two Sample Statistic," Communications in Statistics: Theory and Methods, A5(14), 1393-1407.

Lepage, Yves (1971), "A Combination of Wilcoxon's and Ansari-Bradley's Statistics," Biometrika, 58, 213-217.

Lepage, Yves (1975), "Asymptotically Optimum Rank Tests for Contiguous Location and Scale Alternatives," Communications in Statistics, 4(7), 671-687.

Lepage, Yves (1976), "Asymptotic Power Efficiency for a Location and Scale Problem," Communications in Statistics: Theory and Methods, A5(13), 1257-1274.

Levene, H. (1960), "Robust Tests for Equality of Variances," Contributions to Probability and Statistics, ed. I. Olkin, et. al., Stanford, Stanford University Press, 278-292.

Mann, H. B. and Whitney, D. R. (1947), "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other," Annals of Mathematical Statistics, 18, 50-60.

Mathisen, H. C. (1943), "A Method of Testing the Hypothesis that Two Samples are from the Sa~e Population," Annals of Mathematical Statistics, 14 188-194.

Mehrotra, K. G. and Johnson, Richard A. (1976), "Asymptotic Sufficiency and Asymptotically Most Powerful Tests for the Two Sample Censored Situation," Annals of Statistics, 4, 589-596.

Mood, A. M. (1950), Introduction to the Theory of Statistics, New York, McGraw-Hill.

Moses, L. E. (1963), "Rank Tests for Dispersion," Annals of Mathematical Statistics, 34, 973-983.

Murphy, B. P. (1976), "Comparison of Some Two Sample Means Tests by Simulation," Communications in Statistics: Simulation and Computation, B5(1), 23-32.

Parzen, Emanuel (1961), "Regression Analysis of Continuous Parameter Time Series," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, I, 469-489.

Parzen, Emanuel (1967), Time Series Analysis Papers, San Francisco, Holden-Day.

Parzen, Emanuel (1979), "Nonparametric Statistical Data Modelling," Journal of the American Statistical Association, 74, 105-120.

Parzen, Emanuel (1980), "Density-Quantile Approach to the Nonparametric Two Sample Problem," Stat. 672 notes, Texas A&M University, to be a technical report.

Pearson, E. S. (1931), "The Analysis of Variance in Cases of Non-Normal Variates," Biometrika, 23, 114-133.

Pettitt, A. N. (1976), "A Two Sample Anderson-Darling Rank Statistic," Biometrika, 63, I, 161-168.

Posten, H. O. (1978), "The Robustness of the Two Sample t-Test Over the Pearson System," Journal of Statistical Computation and Simulation, 6, 295-311.

Potthoff, R. F. (1963), "Use of the Wilcoxon Statistic for a Generalized Behrens-Fisher Problem," Annals of Mathematical Statistics, 34, 1596-1599.

Pyke, Ronald and Shorack, Galen R. (1968), "Weak Convergence of a Two Sample Empirical Process and a New Approach to the Chernoff-Savage Theorem," Annals of Mathematical Statistics, 39, 2, 755-771.

Randles, R. H. and Hogg, R. V. (1971), "Certain Uncorrelated and Independent Rank Statistics," Journal of the American Statistical Association, 66, 569-574.

Rosenbaum, S. (1953), "Tables for a Nonparametric Test of Dispersion," Annals of Mathematical Statistics, 24, 663-668.

Rosenbaum, S. (1954), "Tables for a Nonparametric Test of Location," Annals of Mathematical Statistics, 25, 146-150.

Rothenberg, T. J., Fisher, F. M. and Tilanus, C. B. (1964), "A Note on Estimation from a Cauchy Sample," Journal of the American Statistical Association, 59, 460-463.

Savage, I. Richard (1956), "Contributions to the Theory of Rank Order Statistics - the Two Sample Case," Annals of Mathematical Statistics, 27, 590-615.

Scott, David W., Gotto, Antonio M., Cole, James S. and Gorry, G. Anthony (1978), "Plasma Lipids as Collateral Risk Factors in Coronary Artery Disease - A Study of 371 Males with Chest Pain," Journal of Chronic Diseases, 31, 337-345.

Sen, Pranab Kumar (1962), "On Studentized Nonparametric Multi-Sample Location Tests," Annals of the Institute of Statistical Mathematics, 14, 119-131.

Sen, Pranab Kumar (1963), "On Weighted Rank-Sum Tests for Dispersion," Annals of the Institute of Statistical Mathematics, 15, 117-135.

Sen, Pranab Kumar (1966), "On a Distribution Free Method of Estimating Asymptotic Efficiency of a Class of Nonparametric Tests," Annals of Mathematical Statistics, 37, 1759-1770.

Sheffé, Henry (1970), "Practical Solutions to the Behrens-Fisher Problem," Journal of the American Statistical Association, 65, 332, 1501-1508.

Shorack, Galen R. (1969), "Testing and Estimating Ratios of Scale Parameters," Journal of the American Statistical Association, 64, 999-1013.

Šidák, Z. and Vondráček, J. (1957), "A Simple Nonparametric Test of Difference in Location of Two Populations," (Czech.), Aplikace Matematiky, 2, 215-221.

Siegel, Sidney and Tukey, John W. (1960), "A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples," Journal of the American Statistical Association, 55, 429-445.

Smirnov, N. V. (1939), "Estimate of Deviation Between Empirical Distribution Functions in Two Independent Samples," (Russian), Bulletin of Moscow University, 2, No. 2, 3-16.

Steck, G. P., Zimmer, W. J. and Williams, R. E. (1974), "Estimation of Parameters in Acceleration Models," in Proceedings of Annual Reliability and Maintainability Symposium, New York, I.E.E.E.

Sukhatme, Balkrishna V. (1957), "On Certain Two Sample Nonparametric Tests for Variances, Annals of Mathematical Statistics, 28, 188-194.

Switzer, Paul (1976), "Confidence Procedures for Two Sample Problems," Biometrika, 63, 13-25.

Tsao, C. K. (1954), "An Extension of Massey's Distribution of the Maximum Deviation Between Two Sample Cumulative Step Functions," Annals of Mathematical Statistics, 25, 587-592.

Van der Waerden, B. L. (1952), "Order Tests for the Two Sample Problem and Their Power," I, II, III, Indagationes Math., 14, 453-458, 15, 303-310, 15, 80.

Wald, A. and Wolfowitz, J. (1940), "On a Test of Whether Two Samples are From the Same Population," Annals of Mathematical Statistics, 11, 147-162.

Weiss, Lionel (1976), "Two Sample Tests and Tests of Fit," Communications in Statistics: Theory and Methods, A5(13), 1275-1285.

Welsh, B. L. (1937), "The Significance of the Difference Between Two Means When the Population Variances are Equal," Biometrika, 29, 250-262.

Westenberg, J. (1948), "Significance Test for Median and Interquartile Range in Samples From Continuous Populations of Any Form," Proc. Kon. Nederl. Akad. Wet., 51, 252-261.

White, James Michael (1981), "A Quantile Function Approach to the K Sample Quantile Regression Problem," Ph.D. dissertation, Texas A&M University.

Wilcoxon, F. (1945), "Individual Comparisons by Ranking Methods," Biometrics Bulletin, 1, 80-83.

Wilk, M. B. and Gnanadesikan, R. (1968), "Probability Plotting for the Analysis of Data," Biometrika, 55, 1-17.

Zuijlen, M. C. A. Van (1977), Empirical Distributions and Rank Statistics, Amsterdam, Mathematisch Centrum.

DAT
FILM